Journal of Geophysical Research

VOLUME 64 NOVEMBER 1959 NUMBER 11

THE SCIENTIFIC PUBLICATION
OF THE AMERICAN GEOPHYSICAL UNION

Journal of Geophysical Research

An International Scientific Publication

This Journal welcomes original scientific contributions on the physics of the earth and its environment.

Manuscripts should be transmitted to J. A. Peoples, Jr., Geology Department, University of Kansas, Lawrence, Kansas. Authors' institutions, if in the United States or Canada, are requested to pay a publication charge of $15 per page, which, if honored, entitles them to 100 free reprints.

Subscriptions to the *Journal of Geophysical Research* and *Transactions, AGU* are included in membership dues.

Non-member subscriptions, *Journal of Geophysical Research*..$16 for the calendar year 1959, $2 per copy; $20 for the calendar year 1960, $2.25 per copy

Non-member subscriptions, *Transactions, AGU*..................$4 per calendar year, $1.25 per copy

Subscriptions, renewals, and orders for back numbers should be addressed to American Geophysical Union, 1515 Massachusetts Ave., Northwest, Washington 5, D. C. Suggestions to authors are available on request.

Advertising Representative: Howland and Howland, Inc., 114 East 32nd St., New York 16, N. Y.

Beginning with the January 1959 issue (Vol. 64, No. 1) the *Journal of Geophysical Research* is published monthly by the American Geophysical Union, the U. S. National Committee of the International Union of Geodesy and Geophysics organized under the National Academy of Sciences-National Research Council as the U. S. national adhering body. Publication of this journal is supported by the National Science Foundation and the Carnegie Institution of Washington. The new monthly combines the type of scientific material formerly published in the bi-monthly *Transactions, American Geophysical Union*, and the quarterly *Journal of Geophysical Research*. The *Transactions, American Geophysical Union* will continue as a quarterly publication for Union business and items of interest to members of the Union.

Published monthly by the American Geophysical Union from 1407 Sherwood Avenue, Richmond, Virginia.
Second class postage paid at Richmond, Virginia.

Journal of

GEOPHYSICAL RESEARCH

olume 64 November, 1959 No. 11

SYMPOSIUM ON THE EXPLORATION OF SPACE

Introductory Remarks

Robert Jastrow

National Aeronautics and Space Administration
Washington, D. C.

Space physics derives its individuality from s research tools, the satellite and the space robe. The technology of space flight and the pecial requirements of instrumentation for pace experiments separate this field from the raditional areas of geophysics, astronomy, and article physics in which it found its inspiration. he astronomer and the geophysicist who join planning the exploration of the planets find s strong a bond in their common problems as ther has with his colleagues engaged in groundased research.

On April 29–30, 1959, the first nationally ponsored conference devoted to the special roblems of space physics met in Washington nder the auspices of the National Academy of ciences, the National Aeronautics and Space dministration, and the American Physical Soety. Space physics had come into existence vo years before, at a time when only a small art of the scientific community in America was ngaged in the planning and execution of space xperiments. With the establishment of the Naonal Aeronautics and Space Administration, reparations began for an intensified research rogram in the space sciences, and the NASA erefore proposed a symposium on current roblems in space exploration. H. C. Urey had dependently undertaken the organization of a milar conference under the auspices of the ational Academy of Sciences and theAmeran Physical Society; accordingly, the symposium in its final form was organized under the joint sponsorship of the three agencies.

The objectives of the symposium were (1) to awaken the interest of the scientific community in the problems of space research; (2) to present an estimate of our present and future capabilities for space exploration, including the study of the upper atmosphere, the exploration of the moon and near-by planets, and the observation of the sun and stars from orbiting space platforms; (3) to acquaint the experimentalist with existing instrumentation in space physics and to challenge his ingenuity in the construction of new apparatus.

The symposium opened with a session on fields and solid particles in the solar system. In the first paper F. L. Whipple surveys the results of ground-based and satellite meteorite research and describes future possibilities for meteorite instrumentation in space vehicles. It appears that $\approx 10^3$ tons of solid matter fall on the earth daily, that most meteorites are fragile and easily crushed by 0.02 atmosphere, and that all meteorites belong to the solar system. Professor Whipple stresses the potential value of ground-based research for the study of meteors, stating that 80 per cent of the needed information can be obtained by terrestrial observation.

The next two papers, by T. Gold and E. N. Parker, discuss the controversial question of the mechanisms for transport of charged particles from the sun to the earth, and the bearing of these solar corpuscular streams on the fundamental problem of sun-earth relationships. For some time Gold and Parker have been engaged in a running debate on the nature of the transport mechanism, of which a hint remains in the discussion following Parker's paper. Parker bases his model on observations by Biermann that the tails of comets are displaced away from the sun, as though by the pressure of a substantial stream of particles. The observed motions of the comet tails can be explained by a flux of protons with a density of $100/\text{cm}^3$ and a velocity of 500 km/sec. Parker finds that a radial efflux of protons with this density and velocity will be produced by simple hydrodynamic expansion of the solar corona. He therefore calls the radial flow a *solar wind*. In times of enhanced solar activity the solar wind may be strengthened to a density of $10^4/\text{cm}^3$ and a velocity of 2000 km/sec.

According to Gold the radial flow pattern must be occasionally or frequently erased by the eruption of clouds of plasma from the surface of the sun. These plasma clouds are ejected from localized regions côntaining rather high magnetic fields, of the order of 50 gauss. Since the ejected gas clouds are highly ionized they carry the magnetic field with them, but the roots of the field remain fixed in the original region of activity on the surface of the sun. As the gas cloud expands into the interplanetary medium the magnetic field lines are drawn out into loops but are always confined within the limits of the cloud. The more energetic particles emitted in the solar event are in turn confined by the field lines and constrained to travel back and forth within the cloud, in the manner of the Van Allen particles trapped and stored near the earth by the geomagnetic field; Gold calls them solar Van Allen particles.

Gold obtains convincing evidence for the validity of his model by the ingenious association of a variety of observational effects. There is also evidence, however, for a predominantly radial pattern of field and flow lines. It is probable that the final description will have elements of both models. In an effort to discriminate between them, Parker has suggested an interesting experiment requiring the installation of a plasma detector at an appreciable distance from the earth, perhaps at the moon's distance in a lunar satellite. On the occasion of a solar event the enhanced corpuscular flux will be observed first at the space station and later at the earth, and the interval between the two observations will depend on the mechanism of propagation. If the flow is radial and direct, the delay will be a matter of seconds for particles with energies in the million-volt region; but if the particles are magnetically trapped in the expanding gas cloud, which moves at 1000 km/sec, the delay will last several minutes.

The paddle-wheel satellite, Explorer VI, contains a combination of particle detectors and magnetometers which may provide a partial answer to this problem. The payload was designed primarily for other purposes, and the magnetometers are not of the sensitive type required for the study of interplanetary fields Nonetheless, the Explorer VI data obtained during periods of unusual solar activity may provide information of great interest for the understanding of the propagation mechanism.

The most significant single achievement o the IGY was the discovery by Van Allen an his collaborators of a heavily populated zon of geomagnetically trapped particles in th outer atmosphere of the earth. In the fin paper of the session Van Allen summarizes th known properties of the trapped particles an presents new evidence obtained from the Pic neer IV space probe. The Pioneer IV fligh which came after 5 days of unusually intens solar activity, indicated particle fluxes man times greater than those measured by Va Allen's instruments in the Pioneer III rocke In addition, the Pioneer IV data show an ou ward expansion of the trapped particle zone b some 5000 km, and violent fluctuations at tl outer edge of the zone, some 8 or 10 earth rac from the geocenter, suggestive of a disorder domain of mixing between the geomagnetic fie and the incident plasma. Presumably these e fects are produced by the interaction betwe the geomagnetic field and the stream of pr tons emitted from the active regions of t solar surface. Here we are reminded of t remarks by Parker on the instability of the terface between the geomagnetic field and t solar wind. Parker considers that tongues plasma will push through the unstable regi down to a distance of perhaps 5 earth radii fr the center; they cannot penetrate further under ordinary conditions because the geomagnetic pressure will exceed the solar wind pressure at lower altitudes.

The second session opens with a survey of the space program by H. E. Newell. The characteristics of the vehicles in the national space program are described, their capabilities for lunar and planetary research are indicated, and tentative estimates are given for the dates of availability. This paper is directed toward one of the principal objectives of the symposium: it brings home to the participants the imminence of serious lunar and planetary explorations and the great opportunities for research that exist at the moment in the national space program.

The first United States vehicle with a capability for launching payload weights comparable with that of Sputnik III will be the Vega, an Atlas-boosted vehicle which may be available in late 1960 or early 1961, approximately 30 months after the launch date of Sputnik III. The Vega is also the first vehicle with the combined capability in payload weight and guidance needed for major lunar and planetary missions. The capability for soft landings on the moon, and for remote-controlled exploration of the moon's surface, will come only with the Saturn vehicle, a 1.5-million-pound cluster of six or eight smaller engines, which may be ready for these missions in 1963. Manned lunar landings and the return of the landing party to earth will require the Nova, a 6- or 8-million-pound cluster, which will be available in 8 to 10 years. The Nova can also put 75 tons into a 300-mile satellite orbit. This weight will be sufficient to establish a manned space laboratory.

Newell lists the following areas of research in the NASA space sciences program: the atmospheres of the earth and planets, including ionospheres and trapped energetic particles; magnetic, electric, and gravitational fields; solar and galactic observations from orbiting space platforms; and planetary biology. The Goddard Space Flight Center and the Jet Propulsion Laboratory will play important roles in NASA space research, but the major part must be carried out by other research groups, universities, private laboratories, and industry. Proposals from these groups for experimental projects in the above areas are very much desired by the NASA. It may be added that the scientific talent of the country must participate fully in the space sciences program if the United States is to achieve a position of leadership in space research.

The remainder of the session deals with the properties of the moon and planets. The surface features of the moon are discussed by G. P. Kuiper, certain properties of meteorites by H. C. Urey, and the atmospheres of Mars and Venus by G. de Vaucouleurs.

Kuiper suggests that the solid body of the moon was formed by a process of accretion some 5.5 billion years ago. The decay of radioactive elements in the accreted mass caused heating and partial melting in the interior, which led to the appearance of lavas on the lunar surface. Later a cooling phase began, with shrinking and the formation of tension cracks or rills in the surface. Kuiper suggests that the remnant of the old accreted lunar crust may have the texture of a hard rusk, resistant to moderate pressure but yielding under substantial ones, and therefore providing excellent terrain for the first soft landings on the moon.

H. C. Urey discusses the recent evidence for differences in the cosmic-ray ages of the iron and stone meteorites. The irons are dated at some hundreds of millions of years whereas the stones have ages of the order of tens of millions. The apparent youth of the stones may be understood if they have been buried beneath the surface of an object such as the moon and therefore not subjected to cosmic-ray bombardment during the preceding aeons. Accordingly Urey suggests that primary objects of lunar size accumulated about 4.5 aeons ago and that a subsequent disintegration or at least a surface ejection produced the stony meteorites. Among the stony meteorites the chondrites in particular have the character of conglomerates, for whose explanation Urey suggests a violent crushing process in which the original mass was broken into small particles. These were then reaccumulated into a single large mass of substantial strength, which became the immediate parent of the chondrites. If this theory is correct the chondrites that fall on the earth provide us now with samples of lunar material, years before we can hope to determine the composition of the moon's crust by direct exploration. The occasional presence of diamonds in the stony meteorites provides additional evidence for their origin in the moon or a similar object, since the pressures required for the formation of diamonds are to be found only in the interior of an object at least as large as the moon.

The observations by Kozyrev on lunar volcanism are discussed in the second round-table session. D. Alter notes that several independent reports of peculiarities on the floor of the Alphonsus crater had been received at the time of the Kozyrev observation. The consensus of the participants is that some outgassing must indeed have occurred, in the nature of a residual volcanic activity, although probably not the violent eruption usually associated with terrestrial volcanism.

The third session opens with a survey of rocket astronomy by H. Friedman. This new field is making rapid strides in both counter and diffraction grating techniques. An example is the photography of the sun's surface in the Lyman-α line. The Lyman-α photographs show bright plages across the surface which correlate closely with plages photographed simultaneously in calcium K and in Hα, and with the sunspot distributions photographed in white light.

Rocket astronomy has also yielded the first ultraviolet pictures of the night sky. The ultraviolet intensity has a strong maximum over an extended area of the sky, nearly 20° in diameter, around the star Spica in the constellation of Virgo. The ultraviolet brightness of Spica is particularly interesting because this region shows no visible nebulosity when viewed with an optical telescope.

L. Goldberg describes the astronomical observations that will be carried out by satellites in the next few years. Formidable technological difficulties are involved in the attempt to do astronomical research from an orbiting platform, but the National Aeronautics and Space Administration has already undertaken the development of pointing controls and other equipment necessary for the execution of the program. We need only think of the gains in astronomy resulting from the addition of the radio region to the visible spectrum to realize the great potentialities of experiments performed in the extended spectrum accessible to observation outside the atmosphere.

The present plans call for a telescope with a 24-inch objective and associated equipment to be placed in a satellite carrying a total instrument payload of 1 to 2 tons. Orbit stabilization will probably be achieved by gas jets and rotating flywheels. The axis of the telescope will be controlled from the ground. The sun will have high priority among the objectives for investigation from an orbiting platform, with particular emphasis on surface variations in solar activity. The satellite has an advantage over the rocket for these observations because it can monitor the solar surface continuously for an extended period. Monochromatic photography of the surface, eventually with a resolution of 1 second of arc, will greatly advance our knowledge of the structure of the chromosphere.

The program for astronomy from space vehicles also emphasizes the other end of the electromagnetic spectrum, below the 10 Mc/s cutoff imposed on terrestrial radio telescopes by the earth's ionosphere. Radio bursts from the sun and Jupiter are especially important, and again amenable to continuous monitoring by satellite-based antennas.

Satellite investigations of other stars and galaxies are more difficult because the sources are faint and the requirements on pointing control more severe. The first steps in this field will probably consist of extensions of the ultraviolet photographs taken in rockets by Friedman and his collaborators.

The paper by J. W. Townsend is concerned with a second major objective of the symposium, the description of experimental techniques for space research. In the the words of the Chairman, this paper '. . . is particularly important to our symposium because it tells those of us who are not actually working on the construction of space apparatus a great deal about what can be done and how it must be done.' H. E Newell had already outlined the research areas in the NASA science program, and the expected vehicle capabilities and schedules for lunar planetary, and astronomical missions. Townsend provides the details of certain satellite and rocket payloads planned for the second generation of space experiments—payloads of the order of one hundred to several hundred pounds in weight. The devices chosen for his presentation include a number of satellites designed for observations of the upper atmosphere, the ionosphere, and the sun, and two space probes for the measurement of charged particle fluxes and magnetic fields in the outer atmosphere and in the interplanetary medium.

The atmospheric structure satellite is a major project which will include mass spectrometers and several varieties of pressure gage. An attempt will be made to launch this satellite into a polar orbit, permitting direct observation of the anomalous conditions detected in the auroral zone by rockets flown during the IGY. The instruments in the ionosphere satellite include Langmuir probes, ion traps resembling those flown in Sputnik III, an electric field meter to determine the charge on the satellite surface, and an rf probe for the measurement of electron density by the effect of the impedance of an antenna. The primary instrument in the astronomical satellite is an X-ray spectrograph for the measurement of solar radiation in the ultraviolet.

One of the payloads described by Townsend is the 'paddle wheel,' which has now been launched in the Explorer VI satellite at the time of final preparation of these proceedings for publication. The first data received from this satellite indicate that it may return more valuable information on the energetic particle populations, than all previous satellites combined.

In the second space probe payload the principal instrument is a rubidium-vapor magnetometer. This instrument, which is expected to have an accuracy of 10^{-5} gauss, will be the first space probe magnetometer with adequate sensitivity for the investigation of the interplanetary magnetic field. An effort will be made to include a 'solar wind' detector in the magnetometer probe, in spite of the exacting environmental restrictions of the ultrasensitive rubidium-vapor instrument. The interplanetary medium is highly ionized, and the magnetic fields contained in it are therefore frozen into the gas and transported with it; hence the simultaneous detection of fields and particles in one payload provides a much more powerful insight into the structure and dynamics of the medium than either measurement alone.

In the final paper of the symposium R. Jastrow discusses the atmospheres of the moon and the terrestrial planets. The development of the Martian and Venusian atmospheres is first considered very briefly, and the lunar and terrestrial atmospheres are then discussed at greater length.

Jastrow summarizes recent calculations by Herring and Licht in which it is shown that the extent of the lunar atmosphere must be severely reduced by collisions with the protons of the solar wind, if the fluxes in that particle stream have the magnitudes assumed by Parker and others. The trace of lunar atmosphere which exists is probably composed of argon produced by the decay of radioactive potassium in the lunar crust, and possibly other gases produced in such residual volcanic outgassing as has been reported by Kozyrev.

The discussion of the earth's atmosphere considers certain features of the recently obtained data on atmospheric density and temperature. A careful study of satellite orbits has shown a marked correlation between solar activity and satellite drag variations, the latter presumably reflecting corresponding changes in the density of the upper atmosphere. During the investigation of the Sputnik III variations it was found by Jacchia that in two cases the drag increase followed a major solar flare, and could be identified tentatively with the arrival of slow flare particles from the sun. The geomagnetic field channels these solar particles into the auroral regions, where they transfer energy to the atmosphere by collisions. Sputnik III passes through the auroral zone and was apparently affected by the heating and expansion of the atmosphere in that region. The heating effects produced by the channeling of particles into the high latitudes may also account for the anomalously high temperatures and densities that have been derived from rocket measurements in the auroral zone.

Satellite measurements do not show the strong latitude effect that characterizes the rocket data, and this discrepancy between the two sets of measurements has been one of the major problems in the analysis of IGY results. It is expected that the nature of the latitude dependence will be fully understood when data are available from the atmospheric structure satellite.

Journal of Geophysical Research Volume 64, No. 11 November, 1959

Solid Particles in the Solar System

Fred L. Whipple

Smithsonian Astrophysical and Harvard College Observatories
Cambridge, Massachusetts

Introduction

Solid particles in the solar system are here deined as particles greater than molecules and maller than planets; they move about the sun n free orbits and are not physically attached to lanets. Comets, asteroids, and natural satellites night well be included in the subject matter ut will be mentioned only as they are related o smaller particles in the solar system.

Detailed knowledge of the physical structure, omposition and orbital distribution of particles ı the solar system is of particular current inrest because the particles constitute a potenal hazard to satellites, space probes, and space hicles. Both the designers of space equipment nd the operational planners require precise nowledge to protect equipment and operations ainst the effects of meteoritic impact.

Meteoritic material in the solar system can eld fundamental information on the origin and ture of comets and asteroids, which in turn n provide a basis for the development and rification of theories about the evolution of e earth, the planets, and indeed the solar sysn as a whole.

Interactions of meteoritic material with internetary gases, ions, and electrons can yield tistical information relating to the nature and currence of gaseous material in the solar sysn, particularly corpuscular radiation from the 1. Such data may be critical in determining erage conditions of the interplanetary gases. rthermore, interactions of meteoritic material h the moon will aid our understanding of the ar surface and lunar history. Conversely, loration of the lunar surface may tell us ch about meteoritic material in space.

o include a thorough summary of our knowle about solid particles in space would uny lengthen this discussion. Some background erial and a few recent results are presented he next section. Attention centers chiefly on important researches that should be conducted both in space and from the surface of the earth (discussed in subsequent sections). A review of this research brings into focus a striking fact: *an enormous amount of ground-based meteoritic research is urgently required!* Such research will strongly influence the design both of space vehicles and of space experiments and will, at the same time, answer some specific questions about the nature and motions of particles in space.

Ground-based meteoritic studies can provide some 80 per cent of the necessary information about the nature and distribution of particles within 1 astronomical unit of the sun. Extensive studies from space vehicles are in fact needed only to discover unexpected meteor streams or characteristics of meteoritic material in space. Extrapolations beyond the earth's orbit concerning the numbers and orbits of meteoritic particles will very likely become increasingly unreliable with solar distance and probably cannot be entirely satisfactory for space-vehicle design at the distance of Mars. From the ground we seem to have done only about 10 per cent of the pure astronomical research on meteors that could and should be done and less than 1 per cent of the laboratory and range experimental work urgently needed. Experiments with ultravelocity particles, augmented by theoretical developments, will provide fundamental ballistic data from which we can interpret space and ground-based observations and can construct theories of the true nature and motions of solid particles in the solar system. The general subject of meteoritics is in a sorry plight because of lack of interest generally by the scientific community.

Current Meteoritical Results

For general reviews of our present knowledge about solid particles in space and methods of studying them the reader may consult the following: *Whipple and Hawkins* [1959], *Öpik* [1956, 1958], *Whipple* [1958], *Levin* [1956], *Kaiser* [1955], and *Lovell* [1954].

To date the results of meteoritic impact experiments on rockets and satellites have been reported by two groups in the United States and one group in the Soviet Union. In all cases successful experiments have involved ultrasonic impact measurements utilizing piezoelectric detection equipment on surfaces exposed to space, a method first developed by *Bohn and Nadig* [1950]. The results obtained by *Dubin* [1958] at Air Force Cambridge Research Center, by the group at the Naval Research Laboratory including *LaGow, Schaefer, and Schaffert* [1958], and by the Russian experimenter *Nazarova* [1958] agree fairly well regarding the impact frequency on a flat surface exposed randomly in direction from a near-earth satellite. Impacts are recorded usually in the range 0.01 to 0.09 m^{-2} sec^{-1}. Unfortunately the calibration of the equipment used remains so doubtful that interpretation in terms of particle momentum, energy, velocity, or mass distributions is uncertain by a factor of some 10^4 times. The results, therefore, are as yet no more reliable than those obtained by the indirect methods, particularly extrapolation from much larger particles as determined by photographic techniques or from *van de Hulst's* [1947] estimates based upon observations of the zodiacal cloud (probably the most reliable method for total mass influx). It appears likely that the total influx of solids on the earth amounts to a few thousand tons per day, mostly in the form of small particles.

Space vehicles will rarely encounter meteoritic bodies larger than dust. The probabilities of such impacts are fairly well known; some difficulty remains in calibrating the mass corresponding to a meteor of known velocity and luminosity or ion production in the atmosphere. The meteoritic hazards are not particularly great although they must be considered seriously in planetary missions and deep-space probes.

The most precise observational studies of meteors have been made by the two-camera photographic method in the Harvard Meteor Program. The Baker Super-Schmidt meteor cameras have photographed some 6000 meteors simultaneously from two stations, and many results have been obtained by *Whipple and Jacchia* [1957], *McCrosky* [1955*a*, *b*], *Cook and Hughes* [1957] *Whipple and Wright* [1954], *Whipple and Hawkins* [1959], and *Whipple* [1954]. Extensive photographic spectral studies of meteors have been conducted by *Millman* [1955]. The radio observations of meteors can yield radiants and velocities for meteors considerably fainter than those detected by the photographic method, i.e., to the 8th effective visual magnitude or fainter, as compared with the 4th or 5th by photographic detection. Major studies have been conducted by *McKinley* [1951] in Canada. The determination of radiants is due largely to the Manchester, England, group under the direction of *Lovell* [1954]. Particularly active at Manchester have been *Clegg* [1948], *Hawkins* [1956], and *Gill and Davies* [1956]. All the radio meteor work in the United States has been concentrated on problems of radio wave propagation, upper atmospheric research, and communication. *Ellyett*, after his initial studies in England, has with *Keay* [1956], continued radio meteor studies in New Zealand, and *Weiss* [1955], from southern Australia, has also made studies. Considerable activity in radio meteor research is developing in the Soviet Union, but the published results are sparse.

The radio and photographic studies togethe show that to approximately the 8th equivalen visual magnitude all meteorites move in close orbits and belong to the solar system. Th number of hyperbolic meteors, coming from in terstellar space, is at most 1 per cent of th total. The orbits are mostly direct and of lo inclination (about 80 per cent) with no appa ent relationship to the asteroids. The phot graphic evidence clearly indicates that appro mately 90 per cent of all visually observab meteors are of cometary origin [*Whipple*, 1954 No information is yet available on the orb of dust grains except that the dust is chie responsible for the zodiacal cloud and is larg confined to the general region of the eclipt It is safe to assume that near the earth's surfa all meteoritic material travels in the veloc range from 11.3 to 73 km/sec, except concei bly for a small number of particles knock off the moon by meteoritic impact or captu emporarily in orbits about the earth. No proof s yet available that these two latter classes of bjects exist.

The photographic meteor studies by *Jacchia* 1955] and *McCrosky* [1955*a*, *b*] indicate conlusively that ordinary meteors are extremely ragile bodies. McCrosky shows clearly that nany are crushed by dynamical pressures of .02 atmosphere. Unpublished evidence obtained y Cook and Whipple suggests strongly that he densities may be extremely low, possibly elow 0.1 g/cm^3, and that the masses are prob.bly greater than previously expected ($\sim$ 10 g or 0th magnitude). The spectral researches by *Aillman* [1955] indicate that meteoritic maerial is similar in general chemical composition o the heavier elements in the solar system, but o accurate determinations have yet been made f chemical abundances in meteors.

Nothing is known about the density or chemial composition of dust in the solar system, lthough it is presumably cometary in origin. 'he densities are very probably higher than iose believed to occur for the brighter meteors. ine meteoritic dust captured in the atmoshere still remains to be identified positively, though many studies of "meteoritic dust" have een made. Meteorites found on the surface of ne earth have been analyzed extensively by nemical, physical, metallurgical, mineralogical, nd nuclear techniques. The consensus is that ney arose from two or more asteroidal masses hich were shattered into many pieces by many iccessive collisions. That a number of meteites have existed in the solid state for some 5 × 10^9 years is evidenced by the ratio of ^{40} to K^{40} [*Stoenner and Zähringer*, 1958].

Spallation products from cosmic rays in meorites place an upper limit to the total etchg rate of solid surfaces exposed to space. *hipple and Fireman* [1959] have shown on e basis of Fireman's measurements of the arn isotopes, A^{38} (stable) and A^{39} (260-year halfe) in meteorites, that the etching rate does t exceed 2 × 10^{-7} cm/yr for the few meteoes studied to date. If this etching is produced imarily by corpuscular radiation from the n the rate at the earth's distance should not ceed 3 × 10^{-7} cm/yr and may possibly be newhat smaller. The accuracy of this value n still be improved by better determinations of the spallation and shielding effects for cosmic rays of high energies. The etching effect is trivial, in practice, even for optical surfaces exposed several years in space.

A measure of the maximum etching rate on solids in space provides an upper limit to the average amount of corpuscular radiation sent out by the sun. We may assume that the corpuscles are largely protons moving radially from the sun with energies of hundreds or thousands of electron volts. Sputtering occurs on solid surfaces. G. K. Wehner has kindly provided the following tentative estimates of the sputtering rates of protons on iron: 0.5 atom/proton for 10,000 ev or more, 0.2 at 500 ev, and $\sim$0.02 at 100 ev, all at normal incidence. He finds that the rate is greater for moderate angles of incidence, and so I adopt twice these values as applicable to the projected hemisphere of a spherical particle. Also, to apply the results at the earth's distance from the sun, I adopt an etching rate of 3 × 10^{-7} cm/yr. The result is an upper limit to average solar corpuscular radiation as shown in Table 1.

TABLE 1—*Upper limit to corpuscular radiation at earth's solar distance*

Proton energy, ev	Proton velocity, km/sec	Number, cm^3
500	310	290
1000	440	140
5000	980	40

NEEDED EXPERIMENTS FROM SPACE

In the following pages I list some possible types of experiments to be made from space vehicles. Certain critical experiments will be discussed briefly, with comments which are strictly my own opinion and may be subject to considerable alteration as space technology progresses.

Impact Phenomena by Natural Particles Impinging on Exposed Surfaces

The various types of experiments that may be used to study impact phenomena include the following:

Sonic—The utilization, for example, of piezoelectric crystals as detecting mechanisms, the only method successfully used to date in space.

Radiative—the emission of radiation by impacts, measured presumably by photoelectric cells or other highly sensitive radiation-sensing devices with short time constants.

Cratering—the cutting of wires, the puncturing of sheets or thin films, etc., to produce reactions measurable by changes in electrical conductivity or other physical properties.

Large-scale puncturing—the puncturing of thin sheets, in conjunction with mass measures, to determine area and thus density (suggested by McCrosky).

Large-scale cratering—direct study of surfaces, either by recovery from space or by television techniques in space.

The techniques listed do not generally measure the direction of motion of the particles in space unless additional equipment is provided. Furthermore, the experiments do not necessarily separate mass and velocity as independent parameters. Hence, these techniques, if they are to yield adequate information about the solid particles in space, must include the following variable factors: (*a*) orientation of the detector in space (two variables); (*b*) distance and direction of the collector from the earth (three variables); (*c*) position of the earth in its orbit. If the distance from the earth is great, (*b*) and (*c*) can be expressed as: (*d*) radial distance from the sun, and (*e*) distance normal to the ecliptic.

If ground experimentation shows that the methods listed above do not depend upon the same function of mass and velocity for the solid particle, the simultaneous use of two or more of the methods may enable the experimenter to separate mass and velocity as independent parameters. For example, if it were proved that sonic methods measure momentum whereas radiation or cratering methods measure energy, two experiments conducted simultaneously might determine both mass and velocity. The above statements emphasize the vital need for ground-based experimentation with ultravelocity particles if space experiments are to be meaningful.

Total Velocity Vectors of Particles in Space

It is important to note that, if total velocity vectors of particles can be measured in space, the six spatial and time variables listed above, (*a*), (*b*), and (*c*), can be reduced to four. Any experiment giving the distribution function of mass and total velocity vectors near the earth (say, around the earth's equator for a year) will provide knowledge of the orbits about the sun. These orbits, with other available information on the theory of the orbital changes and lifetimes of particles, will give the distribution of particle masses and total velocity vectors everywhere within 1 astronomical unit of the sun. The mass velocity experiment otherwise needs to be utilized as a function of time and place only to determine the location of meteor streams that do not cross the earth's orbit, and for distances well beyond the earth's orbit where other sources of meteoritic particles may be contributing to the background effects. Extrapolations from the earth to the distance of Mars are probably not very reliable.

No method for obtaining the total velocity vector of a meteoritic particle in space has yet been tested. Time differences to microsecond accuracy are required unless the equipment is awkwardly large; 1 cm/μs corresponds to a velocity of 10 km/sec. A few types of possible methods are listed below with critical comments:

Artificial atmosphere contained in space vehicle—Presumably photoelectric techniques could be used in a long tube for the determination of velocity vectors by methods analogous to those used on the surface of the earth. Unpublished calculations by Hawkins indicate tha mercury vapor might not diffuse too rapidly fo a practical experiment.

Mechanical shutter at open end of a tube—Excessive velocities are required for a mechanical shutter, even with a rather long tube and millimeter openings in a 'sausage grinder' typ of shutter.

Rotating drum—This method, suggested b Hawkins, involves an extremely rapid rotatin drum with slits to establish direction. It seem to be limited practically by the strength o materials, with possible vibration difficulties a an additional hazard.

Impact in photographic emulsion—Emulsio techniques are probably unsatisfactory becaus velocity and mass are difficult to separate eve with a number of measuring devices; moreover, the method probably requires recovery.

Grid or thin diaphragm at open end of tube—An impact technique may be feasible in which a diaphragm or grid is placed at the bottom of a tube. Detection would be based on ablation at a thin diaphragm or on electromagnetic effects at a grid. The latter method shows promise worthy of detailed study. Hawkins has suggested the use of grids of various sizes to establish dimension and hence density.

Light or radio reflection from particle as it enters tube—A light or UHF radio source might provide the instant of passage across a tube opening, to the required microsecond accuracy for comparison with the impact instant at the end of the tube. The method shows strong promise and should be explored.

Combined methods—A combination of sonic, radiative, and cratering impact methods may be useful, although the interpretation will be complicated and involve much ground-based experimental calibration.

Erosion on Exposed Surfaces in Space

The possible techniques include the following: (1) conducting strip method, developed at the Naval Research Laboratory; (2) thin radioactive strip, suggested by S. F. Singer; (3) transparency of thin film or window; (4) radiation scattering by etched window.

All these erosion methods show promise of successful operation in space, but, as was indicated in the second section of this paper, the erosion rate appears to be so extremely small that such techniques would require long periods of exposure. If erosion occurs primarily from solar corpuscular radiation, a measurement of erosion rates or of average corpuscular radiation near the earth might suffice for all distances from the sun and near the ecliptic plane. On the other hand, if meteoritic dust is largely responsible, erosion measurements should be made at greater distances than that of the earth from the sun.

Charge on Solid Particles in Space

Such measurements would probably operate by means of a grid and field measuring device backed by an impact diaphragm. They would be particularly useful in the study of solar-terrestrial relationships, with regard to variations arising from position in the solar system, conditions in the interplanetary gas, and corpuscular radiation from the sun. The temperature of the interplanetary gas should be derivable from charge measurements.

Composition of Meteorites

Possible techniques include the following:

Impacting on low-density material—This method probably always requires recovery; it has been attempted with photographic films by the research group at the University of Iowa.

An artificial atmosphere or impacting substance combined with optical spectrograph—This technique might prove extremely difficult but conceivably could be carried out without recovery.

Zodiacal Cloud Measurements

Radiation-sensing elements should be used to determine the intensity and polarization of the zodiacal light as a function of wavelength and of direction in space as seen from the neighborhood of the earth. Measurements from other positions in the solar system might be valuable but are not necessarily critical. It is particularly desirable that such measurements be made at wavelengths less than 3000 A.

Study of Comets

A space probe passing through the neighborhood of an active comet, near the tail or possibly in the coma, would lead to a number of experiments vital to our understanding of the nature of comets. *A fortiori,* a landing on or a close-up view of the cometary nucleus would be invaluable.

Study of Asteroids

A close picture or a landing on an asteroid with proper exploratory equipment would be highly desirable. Research involving landings, by space probes, on near-by comets or asteroids necessitates high space velocities relative to the earth, comparable to those necessary to reach the asteroid belt between Mars and Jupiter.

The Physical Structure of Meteoritic Particles

Whether such studies of the physical structure of particles can be made from space is problematical. It will probably be simpler and more economical to study micrometeorites in the earth's atmosphere.

Exploration of the Lunar Surface

The lunar surface contains a fossil record of meteoritic action on the moon for particles of the entire range from microscopic dust to masses of asteroidal size. Studies of the surface of the moon will provide invaluable information about the impact characteristics of meteoritic material averaged over great intervals of time. More detailed studies will provide additional information about meteoritic material in space; the possible ramifications are so great that no attempt will be made even to list potential methods of attack.

Ground-Based Meteoritical Studies

Although research from space vehicles is of great importance in our study of solid bodies in space, I believe that many of the expected results could be obtained from the ground by an effort small in comparison with that of space research. Whether or not the meteoritical observational programs are indeed fostered from ground-based observatories, it is certain that much of the research from space will be of minimal value unless a great effort is made to develop parallel ballistic studies and experiments on the ground. I consider the experiments listed below of critical importance for the future studies of solid bodies in the solar system, whether these studies are to be conducted from space or from ground-based observatories.

Basic Meteoritic Experiments

Basic ballistics for ultravelocity small particles—Particle masses of importance here range from 10^{-15} to 10^{-8} gram, with velocities in the range from 1 to 73 km/sec. These experiments are fundamental to an understanding of the nature and phenomena of small particles in space.

(*a*) Calibration of sonic impact measuring devices.

(*b*) Impact radiation measurements. These should be conducted in the spectral regions from the ultraviolet through the optical and infrared, possibly to radio wavelengths. No radio radiation has yet been certainly detected from a meteor.

(*c*) Cratering phenomena. Measurement should be made of crater dimensions, directional dependence, and other crater characteristics as functions of density, composition, and structure of both particle and target.

(*d*) Impact explosion products. Velocity vectors and physical state of explosion debris are required as functions of mass, velocity, shape, composition, and density of particle, and composition and density of target. It is particularly important to simulate possible lunar materials as background for lunar exploration.

(*e*) Tests and calibrations of all types of meteoritical detection methods possibly usable in space with all target materials likely to occur on space vehicles.

(*f*) Erosion effects on all materials of interest as produced by fine, ultravelocity particles.

(*g*) Ballistic trajectory measurements in an atmosphere at reduced pressure and in simulated planetary atmospheres. These experiments should include measurements of radiation, ionization, ablation, and drag as functions of mass and velocity.

Ultravelocity ballistics of larger particles—Here the methodology is likely to be somewhat different from and more difficult than for experiments listed above for particles in the range from 10^{-8} gram to practical upper limits in mass and for velocities in the range from 1 to 73 km/sec. These ballistic experiments will serve to calibrate both space and ground-based experiments on solid particles in the solar system and will provide critical theories necessary for the interpretation of these observations. The types of experiments visualized are as follows:

(*a*) All terminal ballistic measures possible such as those listed (*a*) to (*g*) above: Target and projectile materials should be varied among the significant substances as regards composition, density, and physical structure. Both vehicular and lunar surfaces should be simulated.

(*b*) Radiation, ionization, and ablation phenomena in a gaseous medium: Such studies represent the re-entry problem and are presumably well covered in other research programs for minimal meteor velocities. Atmospheric densities should be varied, and the composition should be made to simulate the atmospheres of such planets as Mars and Venus.

(*c*) Studies of ablation products in the atmosphere: Here it would be valuable to study he nature of the ablation products left in the tmosphere, including atoms, ions, molecules, nd dust particles.

Ballistic studies of dust and larger particles an utilize a number of techniques for accelera- on, including acceleration of charged particles ı an electric field, shaped-charge explosions, ght-gas guns, electromagnetic guns, and multi- le-stage rockets. Experiments can be conducted ı evacuated ranges, in the open atmosphere, nd at high altitude from balloons and rockets. hock tubes may be extremely valuable for cer- ain phases of the research.

Very few ballistic experiments of the types sted in this section have been conducted at elocities above 5 km/sec. Since the amount of ork possible in this area is almost infinite and an be expensive, it is recommended that experi- ents be aimed largely at establishing basic hysical theories applicable in wide ranges of rcumstances. Many of the present-day observa- ons would have far more significance were they acked by experiments of the type described oove. The importance of such research can- ot be overemphasized.

adio Meteor Studies

Radio techniques are adequate to determine locity vectors, ionization, and atmospheric ag on meteors corresponding to fainter than e 13th visual equivalent magnitude, i.e., to asses less than 10^{-6} gram at meteoric veloci- es. Only two such projects are under way in e United States, and they are at present verely limited in scope and budget.

Basic radio meteor studies by McKinley and illman in Canada and by Lovell, Hawkins, ıd others of the Manchester Radio Group have ready been mentioned. These studies have led extemely valuable findings concerning the bits and ballistics of naturally occurring me- ors in the earth's atmosphere. Such studies, continued, expanded, and coupled with the ound-based experiments and theoretical ıdies recommended here, can probably answer ost of the questions that could be answered satellite-borne meteor detectors and, in gen- al, answer them more precisely and defini- ely. I am firmly convinced of the truth of e above statement except for the possibility unexpected meteor streams in space and the uncertainty of extrapolation much beyond the range of 1 astronomical unit from the sun.

Radio techniques can determine the spatial-orbit and mass distributions of meteorites striking the earth. With the well determined measures for larger bodies, it will then be possible to calculate orbital changes and lifetimes of particles in space, and predict meteoritic velocity vectors and distributions anywhere within 1 astronomical unit of the sun. As was mentioned before, the extrapolations to greater solar distance will become more and more uncertain with distance.

Furthermore, there are numerous by-products from radio meteor experiments that relate to the geophysics of the high atmosphere, in radio propagation characteristics and in practical forward-scatter communication. Most of the previous radio meteor research in the United States has been aimed at geophysical results and almost none toward the determination of meteoric parameters, phenomena, and physical theory. Exploitation of radio techniques to their limit would reach so nearly the range of particle sizes of interest to space science that other ground-based methods could fill the gap theoretically and leave no major problems on the particle size and velocity distributions for the microscopic dust region.

Studies of Micrometeorites

I define a micrometeorite as a particle so small (or slow) that it radiates away its energy of interaction with the atmosphere without appreciable damage. Included also in the practical studies will be particles produced from larger meteorites by interaction with the atmosphere, and generally characterized as meteoritic dust. The general subject of micrometeorites and meteoritic dust in the atmosphere, on the ground, and in water-laid deposits is in an exceedingly primitive state, although it has almost unlimited potentialities for giving information about the quantity, chemical composition, and physical states of meteoritic dust particles in space. The audience is referred to the book *Meteoritic Dust* by *Buddhue* [1950] and to the recent literature. I am not certain that any airborne or surface microscopic particle has been positively identified as of cosmic origin.

Collection of micrometeorites in deep-sea oozes has been largely promoted by *Pettersson* *and Rotschi* [1950, 1952], particularly during the IGY [*Pettersson,* 1958], but much more effort is required to clarify the identification of individual particles and to prove the cosmic nature of the material. Its rate of sedimentation over many parts of the ocean bottom requires further study, particularly in ancient deposits and in the proposed 'Mohole.'

Collection of meteoritic dust on the surface of the earth and in glacial deposits may or may not yield worth-while results, because of the enormous contamination by terrestrial dust sources. It is quite clear, however, that in the neighborhood of large meteorite falls, particularly of stones, dust collection on the surface of the earth should be intensified. Very fruitful results have been obtained by *Rinehart* [1958] around the Arizona meteorite crater.

Airborne collection techniques, particularly at extremely high altitudes, show the greatest promise of bringing into the laboratory cosmic material that can be identified as such and studied thoroughly.

Optical microscopes have not yet proved successful for identifying and analyzing micrometeorites. Electron-microscope techniques show great promise in this field, but for some time it will probably be necessary to analyze individual particles by such methods as the electron probe X-ray analyzer under development by *Riggs* [1956]. Fireman (unpublished) has applied induced radioactivity techniques to small particles but has yet to identify one certainly of cosmic origin. Michrochemical and spectroscopic methods may eventually be useful if particle-collection techniques can provide sufficient samples. Micrometeorites in the air are reputed by *Bowen* [1953, 1956] to produce sunlight scattering in the twilight sky adequate for measurement. Bowen presents evidence that micrometeorites provide condensation nuclei to set off extraordinarily heavy rainfalls on certain dates of the year. The subject of meteoritic dust and rainfall needs much further investigation.

Analysis of individual particles known to be of cosmic origin will lead to a valuable calibration of zodiacal cloud measurements and to calibration of a number of space-based experiments.

Optical Meteor Studies

Two-camera photographic studies—The Harvard Meteor Program has yielded a number of definitive results with regard to meteoroids large enough to produce visual meteors. As much is now known about orbits of meteors as is known about those of asteroids and comets. Many other facts relating to the physical character of meteorites have been collected. Much more powerful meteor cameras could now be constructed to yield information down to magnitudes of visual equivalent 6 to 8. It is not certain that the gains would justify this extensive and expensive effort. Perhaps a more significant and less expensive undertaking would be the precise photography of very bright meteors (fireballs), as indicated below.

Multicamera fireball photography—This project would involve a network of 9 or more stations separated by distances of approximately 150 km. Each station would cover the entire sky photographically on an automatic basis, so that the velocities, decelerations, luminosities, and orbits of fireballs, detonating bolides, and meteorite falls could be measured. Since there is no strong evidence that the photographed meteors so far include broken-planet debris, the fireball program could bridge the gap between cometary and asteroidal meteoroids. In a few years new meteoritic falls might be discovered in this program, and for the first time precision orbits of meteorites would be known.

The proposed fireball program would produce invaluable information on evolutionary processes, both past and current, among the larger meteoroids, and would provide a better picture of the interplanetary solids in the neighborhood of Mars.

Photographic spectra—*Millman* [1955], who has carried on basic research on the spectra of meteors for more than two decades, has contributed the major known observational results in this area of research. Such spectra could be interpreted much more thoroughly and precisely on the basis of artificial meteor experiments. It is still not possible to estimate quantitatively the original composition of meteoroids because of the inadequacy of theory and calibration under the circumstances of meteroic light production. The use of dispersive equipment on more powerful meteor cameras would contribute appreciably to our knowledge and is urgently recommended.

Ablation and re-entry studies by photographic eteor methods—Two-camera meteor photograhy for the determination of decelerations, luinosities, and spectra has long provided the asic data for a theory of re-entry problems. ractical application of the results has been andicapped by our lack of knowledge of the omposition and physical structure of meteorids. However, these studies have shown that ie meteoroids of cometary origin are extremely agile and probably of very low density. Exeriments at comparable velocities in evacuated anges would do much to clarify the nature of ieteoroids, and it is possible that the subject annot progress far from its present condition rithout such ground-based experiments.

Winds in the upper atmosphere determined y meteor-train observations—Considerable mounts of data have been obtained by twoamera photography of persistent trains left ehind photographic meteors. The method aplies to the range 80- to 120-km altitude, the naximum effect being near 93 km [*Liller and Whipple*, 1954; *Whipple*, 1953].

Air density measures by photographic meteor tudies—This technique has probably outlived ts usefulness in view of the wide variety in the lensity and physical structure of meteoroids ınd the consequent low precision of results.

Some 6000 meteors have been doubly photographed in the Harvard Photographic Meteor Program, and analysis is complete for a large fraction of them. If dispersive equipment could be used on the Baker Super-Schmidt meteor cameras, the spectra of fainter meteors could be obtained and considerable progress made. The next step in direct photography appears to lie either in the direction of much more powerful cameras or toward the network of smaller cameras for fireball photography. Considering the few workers active in the field of meteoritics, I favor the fireball network as leading to the most significant results for the least effort.

Optical Studies on the Zodiacal Cloud

The analyses concerning meteoritic dust by optical studies of the zodiacal cloud and the Fraunhofer corona, initiated by *van de Hulst* [1947] and *Allen* [1947], are being continued by *Elsasser* [1958], *Haug* [1958], and others. Much more study, at all the optical wavelengths available, would be fruitful. At extremely high altitudes these investigations might lead to a definitive determination of variations in brightness with seasonal or with solar activity. Polarization is a most important factor. In combination with wavelength variations, polarization data might establish more about the physical-chemical structure of meteoritic dust in space and luminosity produced by electron scattering.

Much more experimental and theoretical research is needed to establish the scattering characteristic of the meteoroid dust balls and to interpret the zodiacal cloud observations in terms of orbital distribution for the particles.

Meteorites

Only a beginning, hardly more than a taxonomic approach, has been made in the basic study of meteorites by classical methods, including metallurgical, mineralogical, physical, optical, and chemical techniques. It is highly desirable that all these methods, particularly the precision chemical analysis, be applied systematically and thoroughly to the meteorites in our various collections. As *Urey* [1955, 1956, 1957] and others have shown, meteorites provide a key to our knowledge about the formation of the asteroids and therefore to the planetary system in general. Before using space exploration for the solution of those problems we must make some attempt to exhaust the sound and relatively inexpensive method of study now available for application to meteorites. In this area I cannot overemphasize the significance of a relatively small basic research program to answer fundamental questions about solid material in space and the primary cosmological problems.

Modern techniques in radioactivity and nuclear physics have shown the way to enormous progress in the study of meteorites. *Bauer's* [1947] evidence for the effects of cosmic rays in producing helium in meteorites has been amply confirmed by the work of *Paneth, Reasbeck, and Mayne* [1951], *Fireman and Schwarzer* [1957], *Stoenner and Zähringer* [1958], and others. *Fireman's* [1958*a*] work has already led to a measure of the upper limit of erosion on meteorites in space [*Whipple and Fireman*, 1959] and has provided measures of the ablation effects of atmospheric passage on two large iron meteorites [*Fireman,* 1958*b*, 1959]. These methods, if thoroughly exploited, can lead to many unanticipated results with regard to the effects of cosmic rays on bodies in deep space, the nature and origin of meteorites, and other problems of great practical and theoretical interest. This fresh, new method of attack appears to be of extreme value and potentiality. Its exploitation should be accelerated.

Studies of Meteorite Craters

Very little progress has been made on the subject of great impact crater formation on the earth, except for the final acceptance of *Barringer's* [1905] interpretation of the Arizona meteor crater, and the general recognition by geologists of a few structures of meteoritic impact origin. Recently *Beals* [1957] has done outstanding work in identifying geologically covered and eroded craters in Canada, and has added materially to our understanding of the nature of such structures. Research on these huge craters, however, is still in its infancy and should be encouraged. Not only do such studies clarify the problems of crater formation by huge explosions but they also give us ancient historical information of geologic value as well as an insight into the occurrence of large, solid particles in space during geological history.

It seems premature to make intensive studies of craters formed on the moon before we have thoroughly investigated those available locally on the earth. Thus far, only little research has been done on the great craters in the United States.

Current Research by Astrophysical Methods

Increased observational and theoretical studies of comets are clearly indicated if we are to understand the occurrence of small particles in space, since the photographic meteor studies show clearly that comets are the major contributors to meteoritic dust. Much valuable research on comets can be conducted with the aid of modern auxiliary equipment in conjunction with astronomical telescopes.

Studies of Asteroids by Astrophysical Methods

Most of the physical studies of asteroids since 1950 have been conducted by *Kuiper and others* [1958]. Continued studies of these bodies should provide results of high significance to space probes, particularly in the ranges of the solar system between Mars and Jupiter.

Perturbation Studies of the Motions of Astronomical Bodies

Many more theoretical studies are required to understand the origin of comets, the variations in their orbits produced by the perturbations of the planets, and the changes in the orbits of smaller bodies caused by perturbations, by solar radiation, and by interplanetary gases. Many of the developments required for the solution of the problems of comets, asteroids, and interplanetary particles will lead to techniques and results significant in space research. It is to be hoped that modern methods of analysis and computing can be used much more extensively in these various basic problems of celestial mechanics than they have in the past.

I am especially grateful to G. S. Hawkins, L. G. Jacchia, and R. E. McCrosky for discussions and suggestions that have added materially to this paper. G. K. Wehner has been most generous in permitting me to publish his tentative results on the sputtering rates of protons on iron.

References

Allen, C. W., The spectrum of the corona at the eclipse of 1940 October 1, *Monthly Notices Roy. Astron. Soc., 106,* 137–150, 1947.

Barringer, D. M., Coon Mountain and its crater, *Proc. Acad. Nat. Sci. Phila., 57,* 861–886, 1905.

Bauer, C. A., Production of helium in meteorites by cosmic radiation, *Phys. Rev., 72,* 354, 1947.

Beals, C. S., A probable meteorite crater of great age, *Sky and Telescope, 16,* 526–528, 1957.

Bohn, J. L., and F. H. Nadig, Researches in the physical properties of the upper atmosphere with special emphasis on acoustical studies with V-2 rockets, *Research Inst. Temple Univ., Rept. 8,* 1–26, 1950.

Bowen, E. G., The influence of meteoritic dust on rainfall, *Australian J. Phys., 6,* 490–497, 1953.

Bowen, E. G., The relation between rainfall and meteor showers, *J. Meteorol., 13,* 142–151, 1956.

Buddhue, J. D., *Meteoritic Dust,* Univ. New Mex. Publ. Meteoritics, no. 2, Albuquerque, 1950.

Clegg, J. A., Determination of meteor radiants by observation of radio echoes from meteor trails, *Phil. Mag., 39,* 577–594, 1948.

Cook, A. F., and R. F. Hughes, A reduction method for the motions of persistent meteor trains, *Smithsonian Contribs. Astrophys., 1,* 225-238, 1957.

ubin, M., Direct measurements of meteoritic dust using rockets and satellites, Tenth General Assembly of International Astronomical Union, Moscow, August 1958.

llyett, C. D., and C. S. L. Keay, Radio echo observations of meteor activity in the southern hemisphere, *Australian J. Phys., 9,* 471–480, 1956.

lsasser, H., Interplanetare materie, *Mitt. astron. Ges. Hamburg,* 1957 II, 61–88, 1958.

ireman, E. L., Argon-39 in the Sikhote-Alin fall, *Nature, 181,* 1613, 1958*a*.

ireman, E. L., Distribution of helium-3 in the Carbo meteorite, *Nature, 181,* 1725, 1958*b*.

ireman, E. L., The distribution of helium-3 in the Grant meteorite and a determination of the original mass, *Planetary and Space Science, 1,* 66–70, 1959.

ireman, E. L., and D. Schwarzer, Measurement of Li^6, He^3, and H^3 in meteorites and its relation to cosmic radiation, *Geochim. et Cosmochim. Acta, 11,* 252–262, 1957.

ill, J. C., and J. G. Davies, A radio echo method of meteor orbit determination, *Monthly Notices Roy. Astron. Soc., 116,* 105–113, 1956.

aug, U., Über die Haufigkeitsverteilung der Bahnelemente bei den interplanetaren Staubteilchen, *Z. Astrophys., 44,* 71–97, 1958.

awkins, G. S., Radar echoes from meteor trails under conditions of severe diffusion, *Proc. IRE, 44,* 1192, 1956.

awkins, G. S., and R. B. Southworth, The statistics of meteors in the earth's atmosphere, *Smithsonian Contribs. Astrophys., 2,* 349–365, 1958.

ulst, H. C. van de, Zodiacal light in the solar corona, *Astrophys. J., 105,* 471-488, 1947.

acchia, L. G., The physical theory of meteors, VIII, Fragmentation as a cause of the faint-meteor anomaly, *Astrophys. J., 121,* 521–527, 1955.

aiser, T. R., Editor, *Meteors,* Pergamon Press, London, 204 pp., 1955.

uiper, G. P., Y. Fujita, T. Gehrels, I. Groenevald, J. Kent, G. van Biesbroeck, and C. J. van Houten, *Astrophys. J.,* Suppl. 3, no. 32, 289–428, 1958.

aGow, H. E., D. H. Schaefer, and J. C. Schaffert, Micrometeorite impact measurements on a 20-inch diameter sphere at 700 to 2500 kilometers altitude, CSAGI Meeting of the IGY, Moscow, August 1958.

evin, B. J., *Physical Theory of Meteors and Meteoritic Matter in the Solar System,* Academy of Sciences, Moscow, U.S.S.R., 293 pp., 1956.

iller, W., and F. L. Whipple, High-altitude winds by meteor-train photography, in *Rocket Exploration of the Upper Atmosphere,* R. L. Boyd and M. J. Seaton, Editors, Pergamon Press, London, pp. 112–130, 1954.

ovell, A. C. B., *Meteor Astronomy,* Clarendon Press, Oxford, 463 pp., 1954.

McCrosky, R. E., Physical and statistical studies of meteor fragmentation, Thesis, Harvard University, 1955*a*.

McCrosky, R. E., Fragmentation of faint meteors (Abstract), *Astron. J., 60,* 170, 1955*b*.

McCrosky, R. E., and A. Posen, The new photographic meteor showers, *Astron. J., 64,* 25–27, 1959.

McKinley, D. W. R., Meteor velocities determined by radio observations, *Astrophys. J., 113,* 225–267, 1951.

Millman, P. M., Meteor news: photographic meteor spectra (Appendix 3), *J. Roy. Astron. Soc. Can., 49,* 169–172, 1955.

Nazarova, T. N., Rocket and satellite investigation of meteors, Tenth General Assembly of the International Astronomical Union, Moscow, August 1958.

Öpik, E. J., Interplanetary dust and terrestrial accretion of meteoritic matter, *Irish Astron. J., 4,* 84–135, 1956.

Öpik, E. J., *Physics of Meteor Flight in the Atmosphere,* Interscience Publishers, New York, 174 pp., 1958.

Paneth, F. A., P. Reasbeck, and K. I. Mayne, Helium 3 content and age of meteorites, *Geochim. et Cosmochim. Acta, 2,* 300–303, 1951.

Pettersson, H., Rate of accretion of cosmic dust on the earth, *Nature, 181,* 330, 1958.

Pettersson, H., and H. Rotschi, Nickel content of deep-sea deposits, *Nature, 166,* 308, 1950.

Pettersson, H., and H. Rotschi, The nickel content of deep-sea deposits, *Geochim. et Cosmochim. Acta, 2,* 81–90, 1952.

Riggs, F. B., Jr., Meteoritic research at Harvard, *Harvard Alumni Bull.,* pp. 15-19, October, 1956.

Rinehart, J. S., Distribution of meteoritic debris about the Arizona Meteorite Crater, *Smithsonian Contribs. Astrophys., 2,* 145–160, 1958.

Stoenner, R. W., and J. Zähringer, Potassium-argon age of meteorites, *Geochim. et Cosmochim. Acta, 15,* 40–50, 1958.

Urey, H. C., On the origin of tektites, *Proc. Natl. Acad. Sci., U. S., 41,* 27–31, 1955.

Urey, H. C., Diamonds, meteorites, and the origin of the solar system, *Astrophys. J., 124,* 623–637, 1956.

Urey, H. C., Origin of tektites, *Nature, 179,* 556–557, 1957.

Weiss, A. A., Radio echo observations of meteors in the southern hemisphere, *Australian J. Phys., 8,* 148-166, 1955.

Whipple, F. L., Winds in the upper atmosphere by meteor-train photography, *J. Meteorol., 10,* 390–392, 1953.

Whipple, F. L., Photographic meteor orbits and their distribution in space, *Astron. J., 59,* 201–217, 1954.

Whipple, F. L., The meteoritic risk to space vehicles, *Proc. VIII Intern. Astronautical Congr., Barcelona, 1957,* Springer-Verlag, Vienna, 418–428, 1958.

Whipple, F. L., and E. L. Fireman, Calculation of erosion in space from the cosmic-ray exposure age of meteorites, *Nature, 183,* 1315, 1959.

Whipple, F. L., and G. S. Hawkins, Meteors, *Handbuch der Physik,* S. Flügge, Editor, Springer-Verlag, Berlin, *52,* 519–564, 1959.

Whipple, F. L., and L. G. Jacchia, The orbits of 308 meteors photographed with Super-Schmidt cameras (Abstract), *Astron. J., 62,* 37, 1957.

Whipple, F. L., and F. W. Wright, Meteor stream-widths and radiant deviations, *Monthly Notices Roy. Astron. Soc., 114,* 229–231, 1954.

Discussion

Question: You mentioned the desirability of a further study of meteor craters on the earth. Several years ago, an article was published on the sighting of meteor craters from the air. I wonder if such a survey has been continued.

Mr. Whipple: I don't think that has been done consistently. You must in any case follow up with studies on the ground by drilling, measuring gravity anomalies, magnetic anomalies, and actually getting into the rubble defining the crater.

Question: Dr. Whipple, you mentioned that the two-camera technique was perhaps near the end of its usefulness. Still, the two-camera data which you obtained are the only quantitative data we have on meteor entry into the atmosphere. Therefore, I am interested in your statement that you have measured 6000 of these in the last few years. Are these data being reduced, and will they be available?

Mr. Whipple: There are three programs in that area. Dr. Jacchia has done about 500 of the longest and most precise tracks. Dr. Hawkins has studied 400 randomly selected cases, mostly shorter ones. And McCrosky has studied 2500, taking everything that was measurable out of the 6000. Those data are being prepared for publication now.

Plasma and Magnetic Fields in the Solar System

THOMAS GOLD

Harvard College Observatory
Cambridge, Massachusetts

I am planning to talk about the interplanetary as in the inner part of the solar system, in hich we live. I should like to discuss the cir-umstances that we shall find when we send p suitable instruments to investigate that gas.

The subject is the counterpart to meteorology n the earth; we shall be concerned with the ermanent and the variable features in the gase-us content of the inner part of the solar sys-em. We first ask, what will be the experiments quivalent to the meteorological measurements f temperature, pressure, and wind? What are he quantities we should be interested in, and hat are the orders of magnitude that we now xpect we shall find?

Of course we know little at present. We have nly some very indirect methods of inference, nd it is essential to carry out the space-vehicle xperiments that are now in prospect. Neverthe-ess, it is best to state what can be inferred rom the meager data available, because this tatement will suggest suitable experimentation.

We know that the space between the earth nd the sun is not empty. Apart from the solid odies and the dust particles about which we ave just heard, there is some gas of very un-ertain density. Frequently, or some of the ime—I am not sure which is more correct—the ensity is between 100 and 1000 particles per ubic centimeter. It may be well above that igure on occasions; it may also be below.

This material probably consists predominantly f protons and electrons. Since it is almost cer-ainly highly ionized, magnetic fields will move with the gas. If this material is moving, we must xpect that magnetic configurations will also move in the solar system. Thus, we must completely give up the picture of static fields extending out to large distances from the sun and the earth. At some distance from the sun and at some distance from the earth the motion of gas masses will completely dominate and the magnetic fields will be convected around by that motion. Hence the intensity and direction of the interplanetary magnetic field will be one type of measurement that we shall be very much concerned with in the exploration of the solar system.

In confirmation of this suggestion we note that the time constants for dissipative decay are certainly thousands or hundreds of thousands of years. The interplanetary gas velocity that we estimate from the correlation between solar events and terrestrial phenomena is generally of the order of 1000 km/sec. Therefore the material takes a day or two to pass through an astronomical unit, an interval in which the dissipative decay of magnetic fields is completely negligible.

We think that there are two types of sources on the sun that send material to us. From solar observation it is certain that one source resides in the chromosphere, close to the sun's surface, from which on occasions a terrific blast of material comes out and hits us 1, 2, or 3 days later.

There is probably another source higher up, consisting of coronal material. We shall hear about that from Professor Parker later.

The chromospheric outbursts are the sudden events, the flares on the sun and the magnetic storm phenomena on the earth starting sometime later. For this kind of event we have a fairly reliable way of estimating the densities involved. The observed magnetic storms require particle densities in excess of 1000/cm^3, with velocities between 500 and 2000 km/sec.

It is also clear from observations of magnetic storms that solar outbursts of this type do not diffuse in the process of traveling from the sun to the earth, because the onset of the storm at the earth frequently occurs very sharply, in a matter of 1 minute, although the total transit time from the sun has been perhaps 2 days. The streaming out of such material from the sun, then, does not proceed into a vacuum, for the thermal diffusion of an expanding cloud would have assured a gentle rise of the magnetic storm phenomenon over many hours; rather, it is proceeding with a sharp wave front. This front cannot, I think, be interpreted except as the shock wave between the material ejected from the sun and the material that is normally in the space between sun and earth. A shock front would collimate the material in time and in space so that a sudden event could arise here.

What are the consequences of the magnetic fields that we would expect such an outburst to carry along? I believe that problem was first discussed by Professor Morrison in 1954 or 1955, when he considered the effects that would be produced by magnetized clouds coming from the sun. He showed that an expanding system of gas carrying magnetic fields would cause a decrease in the cosmic-ray flux.

It is not difficult to see that an expanding system containing a magnetic field will deflect the cosmic-ray trajectories and reduce the flux in the center of the region.

Recently we have been able to form a somewhat clearer picture of the types of magnetic configurations that must result from solar outbursts. If an outburst comes from a solar region in which magnetic fields of the order of 50 or 100 gauss are common—and we think that this is the situation in disturbed regions of the chromosphere—then, when this gas is drawn out and ejected, it must remain magnetically connected to the roots of the field in the sun. Whatever the initial shape of the field at the surface of the sun, if the field protrudes into the chromosphere from the photosphere, as it normally does, its lines of force will be drawn out and will continue to follow with the material. The material might be disturbed in a variety of ways, but the lines of force from the ejected material will connect back to the sun. A solar outburst must therefore take the form of a growth outward of such magnetized region, as suggested in Figure 1.

What will be the consequence of these field configurations for cosmic rays that normally enter from all directions? If they were completely prevented from entering the magnetic field, there would be contained within this region only those cosmic rays that were within the region before it started to expand from the sun. There it had a very small volume and would have contained very few cosmic rays. The cosmic-ray flux, therefore, would be diminished enormously. In fact, the holding out of the cosmic rays will not be perfect if the fields are not strong enough, and there may be only a moderate diminution in the cosmic-ray intensity (Fig. 2).

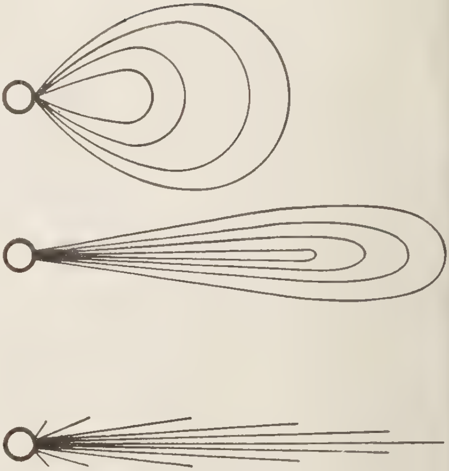


Fig. 1—Suggested shapes of magnetic field lines resulting from a chromospheric outburst. At left and center are shown possible shapes in a plane containing some field lines; at right is a projection on a normal plane. The distances indicated are such that effects of solar rotation would not be significant.

One would therefore expect an outburst resulting in such a magnetic configuration to produce on the earth a magnetic storm phenomenon when the gas hits the earth and at the same time to cause a decrease in the cosmic rays as a consequence of the magnetic configurations sweeping over the earth. This is the very commonly observed correlation between the so-called Forbush decrease in the cosmic radiation, which is a diminution of a few per cent in the general flux, and magnetic storms on the earth.

Can we say anything about the strength of the magnetic field from the size of the observed Forbush decreases? We can make an estimate

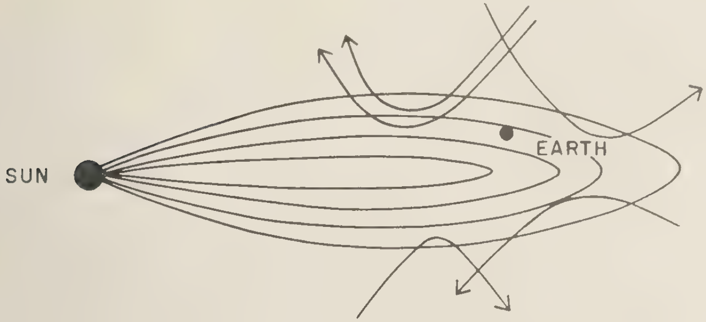


Fig. 2—A deflection of cosmic rays by a magnetic field carried with the solar plasma.

r an occasion when the field was apparently
ery strong, namely, a Forbush decrease in
hich the cosmic-ray intensity diminished with
usual rapidity. The information was obtained
om Professor Rossi's group at MIT, which
bserved a Forbush decrease occurring within
hour, the most rapid decrease that has been
bserved, although a few hours is common.

A small percentage of the particles of mag-
etic rigidity of 10^{10} ev/c were held off in this
ecrease. The speed of the advancing front was
idently of the order of 1000 km/sec reckoned
rom the delay time between the solar event and
e Forbush decrease. The distance moved in 1
our was therefore about 10^{11} cm, and this dis-
ance must be of the order of a radius of gyra-
on of the fast particles that are held off. From
is distance and the magnetic rigidity we can
onclude that the field expanding over the earth
t the time had a strength of the order of 10^{-4}
auss. This value is probably high, because
ost Forbush decreases are less pronounced
an the present example; possibly 10^{-5} gauss
ould be more typical.

If cosmic rays can be excluded from a certain region, it follows that they can also be confined in that region if they are produced inside it. Particles could be generated at the sun and caught in such orbits as those shown in Figure 3. These orbits are similar to the trapped orbits of the Van Allen layer.

In such a model it is possible to have particles moving through the vicinity of the earth on steep-pitched spiral orbits in which they may be stored for a long time.

What observational evidence have we for the existence of solar Van Allen particles? From observation, there are three energy ranges for particles of solar origin. In each of them there is some suggestion of particle storage for appreciable periods, in agreement with the prediction of our model.

First, there are a few occasions, perhaps five in all, on which particles of *cosmic-ray energy* are known to have come from a large solar flare. In at least one of the five, the flare of February 23, 1956, which was also the best documented, particles with momentum of 3 × 10^{10}

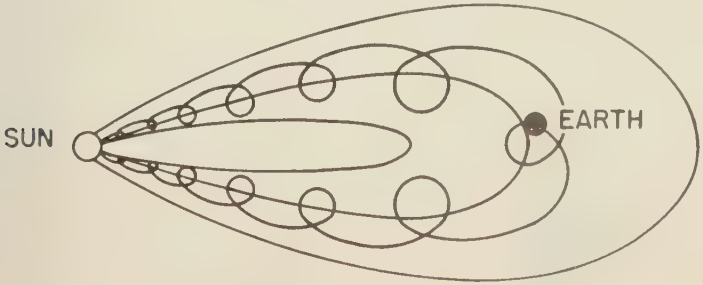


Fig. 3—Captured solar cosmic rays.

ev/c arrived directly, followed by particles of somewhat lower energy, with momentum of 10^{10} ev/c. The less energetic particles continued to arrive for some hours after the flare, and apparently from all directions in the solar system. If the sustained arrival and isotopic incidence of these particles are taken as evidence of magnetic storage, we can say that particles with momenta of 10^{10} ev/c must have moved through a magnetic field sufficient to cause them to do a few turns between sun and earth, because orbits that do not make a single turn would be lost out of the region. In fact, storage was seen for a period of some hours; we can therefore say that the stored particles must have done several turns. For the purpose of estimating the field strength required, I have set the number of turns at 4, although it might be a little larger.

In that case, for the rigidities observed, it is necessary to have about 10^{-5} gauss in the region:

$$H = \frac{E}{300\rho}$$

$$= \frac{10^{10}}{300 \times (1.5 \times 10^{13})/4} \approx 10^{-5} \text{ gauss}$$

Second, solar particles are also observed in an energy range corresponding to a penetration into the atmosphere to an altitude of 50 km, or an energy of ≈100 Mev. This group of particles has been investigated only very recently. We have some knowledge of them from experiments by Anderson and others, and from observations of radio effects, of which I have chosen those that *Bailey* [1957] has collected to mention here.

It appears that particle fluxes in this range can be recognized by their ionospheric effects. I will not go into details of these effects but merely say that it is beyond any reasonable doubt that their causes are particles that have come from large solar flares with time delays of the order of an hour and that they show storage for a matter of many days, certainly 4 and possibly 7 or 8 days on occasion. These particles of subcosmic-ray energies, by showing the very long storage, make it possible to observe a propagation to the earth in two different ways. The particles can propagate at their own speed on their spiral orbits from a disturbance on the sun to the earth, provided that the magnetic field configuration in the intervening space is suitable. They will have very small radii of gyration, so that unless the field is exactly suitable they will not hit the earth from a particular disturbed region on the sun. At lower particle energies, the probability for reaching the earth directly from a particular disturbance on the sun diminishes. Another method of propagation then becomes possible: on or near the sun a certain storage region of plasma and magnetic field is filled with these low-energy particles, and a part of that storage region expands with the motion of the gas so as eventually to include the earth.

Thus we have the possibility of an indirect method of propagation, in which the particles do not come to us at their own travel speed but rather at a speed dictated by the comparatively slow expansion of the gas moving and carrying the magnetic field with it. Such particles are merely stored and gain access to the earth later, depending on the travel time of the gas.

This indirect method of propagation will become more and more probable as we go to lower particle energies. On general grounds one would expect that the sun is more frequently able to make the lower-energy particles than the higher ones.

The high-latitude ionospheric effects discussed above are caused by protons with energies between 10 and 100 Mev. In that energy range we expect that we may be able to see both the direct and the indirect method of propagation.

Figure 4 shows Bailey's collection of data on the occasion of the great flare of February 23. The heavier lines show the ionospheric propagation effects that set in sharply at the time of the flare and persist for several days. Superimposed on the anomalous ionospheric absorption is a diurnal effect not relevant to the present argument.

The same kind of phenomenon has been seen many times since, but without the simultaneous occurrence of energetic cosmic rays. In other words, the sun does indeed make these low-energy particles more often than it makes particles with energies in the bev range.

Figure 5 (Bailey, private communication)

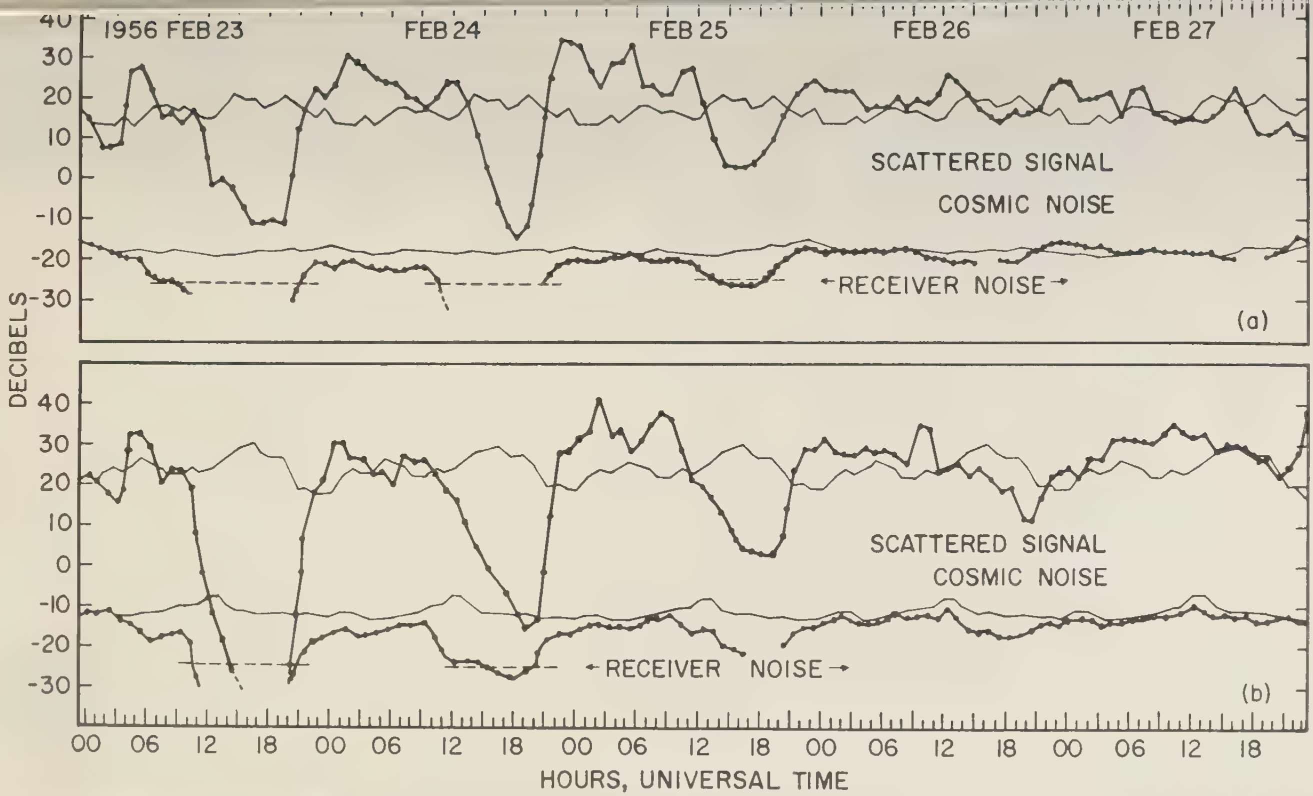


FIG. 4—Ionospheric absorption possibly produced by solar protons emitted in the flare of February 23, 1956. The records present observed variations with time of signal intensity and background cosmic noise, received at the same frequency at the same antenna. The normal variations are indicated by the thin lines. Vertical arrows indicate local apparent noon at path midpoint. (*a*) Observations at Søndre Strømfjord, Greenland, of signals from Thule, Greenland, and background cosmic noise, at 31.5 Mc/s. (*b*) Observations at Søndre Strømfjord of signals from Goose Bay, Labrador, and background cosmic noise, at 32.2 Mc/s [*Bailey,* 1957].

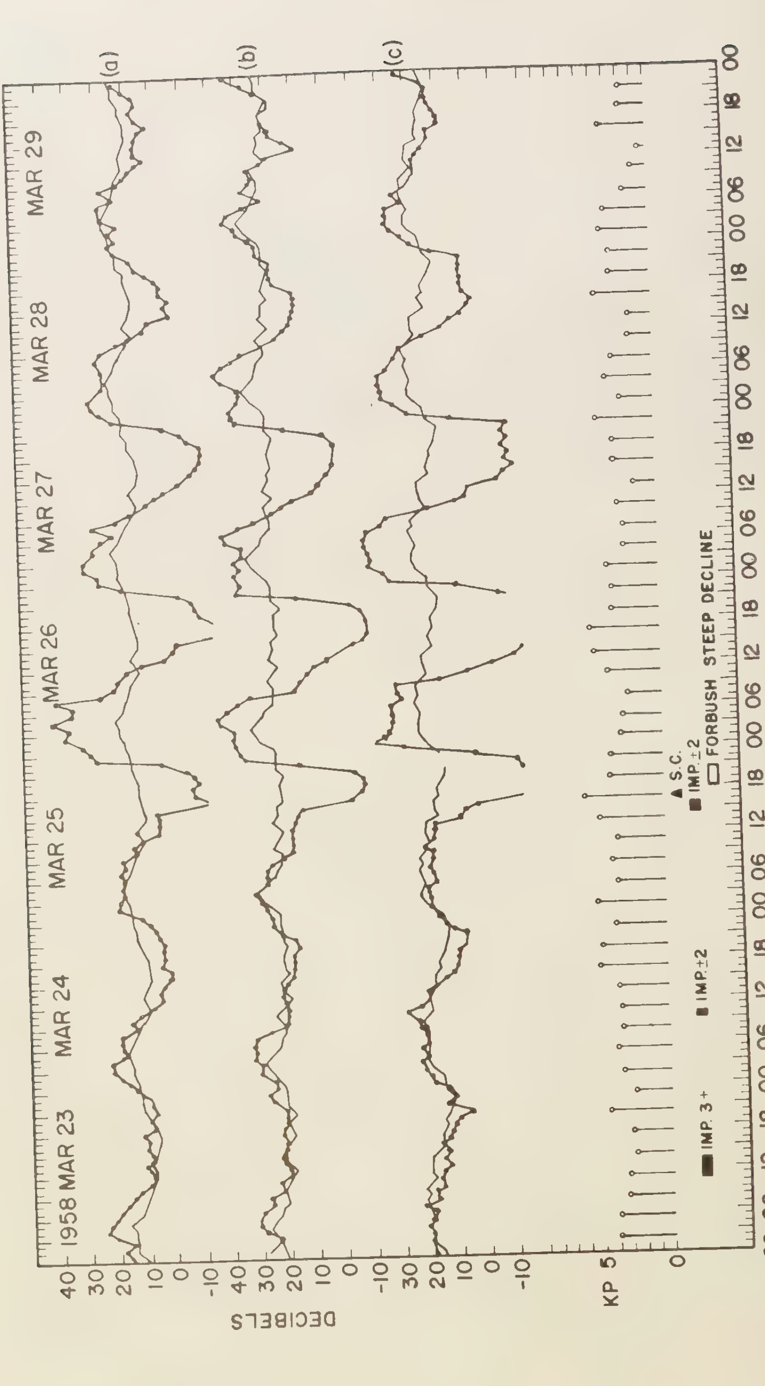


Fig. 5—Ionospheric effect produced by low-energy particles emitted in a major flare on March 22, 1958. The anomalous ionospheric absorption occurs March 25 simultaneously with the onset of a magnetic storm (SC) and a Forbush decrease (Bailey, private communication).

provides further evidence for anomalous absorption produced by the low-energy particles. The mark at 1200 UT on March 22 refers to a flare of importance 3+ on the sun—a very large flare. It was followed 2 days later, which appears to be a very common delay time, by a magnetic storm starting at the triangle *SC* and at the same time by a Forbush decrease certainly associated with the other events.

Thus there occurred on the occasion of this flare such an outburst as is very common, of a magnetized region leading to a Forbush decrease, and the impact of the gas in this region leading to a magnetic storm on the earth.

But nearly simultaneously with these two events there occurred an increase in the flux of the subcosmic-ray particles. Clearly we have here a case of indirect propagation and a demonstration that it was the same magnetic region between the sun and the earth that was responsible both for the exclusion of the ordinary cosmic rays, namely the Forbush decline, and for the inclusion of solar particles of low energy in great numbers giving rise to the same ionospheric event that we saw before. These particles were presumably stored in the magnetic region, had been stored in it for the 2 days preceding the phenomenon, and continued to be stored for some 5 days afterward.

We have here an event of just the nature that our model would suggest, namely, that the same magnetic region is capable of both including and excluding particles, that the particles of solar origin are captured within it and the external cosmic rays are excluded from it.

If we consider particles of still lower energy, it becomes improbable that they will ever get here by the direct method of propagation and more probable that they will arrive only by the indirect method. As they are likely to be produced more often, this should be a common occurrence.

Outbursts on the sun resulting in magnetic storms on the earth can be expected commonly to result in storage of these low-energy particles, whose energy may nevertheless be much in excess of the thermal gas energies that come in the outburst.

Such events will lead to the arrival here of particles not energetic enough to cause the ionospheric event reported by Bailey, and therefore much harder to detect.

Since the discovery of the radiation zones by Van Allen, it is tempting to speculate that the low-energy particles are often fed into the storage region around the earth by the same processes that stir up the magnetic field of the earth and make a magnetic storm.

Perhaps we are observing the effects of a continuous spectrum of solar particle energies, and we are finding that the most energetic ones can come to us only by the direct channel, and the less energetic only indirectly. The particles of very low energies cannot enter the earth's field by any means, but we can see them get in when the field is disturbed by plasma motion, that is to say, by magnetic storm disturbances. Thus we span, possibly with the same intrinsic solar spectrum, all the phenomena from the cosmic-ray outbursts on the sun to the supply of the Van Allen layer by the particles producing the magnetic storms. Such a point of view requires that the particles injected at the outer edge of the Van Allen zone be able to diffuse thoroughly through the zone. If the particles are supplied in an outer arc, they must later be able to appear on inner arcs. The lifetimes of these particles, once they are in captured orbits, may be very long, and therefore the rate at which they have to be diffused is not necessarily very high, but they must be able to get around.

The region in the vicinity of the earth in which the earth's magnetic field dominates all dynamical processes might be called the 'magnetosphere.' The laws of motion of ionized material in the magnetosphere are greatly affected by the fact that the entire magnetospheric conducting region is separated from the conducting earth by the insulating sheath of the nonionized dense atmosphere.

Therefore we have no right to assume an intrinsic stability of the material on one particular tube of force or on one particular surface of revolution of a tube of force around the axis as we would if those lines of force were anchored in a conductor. Because of the intervention of the insulating atmosphere, however thin, the conducting material on one line of force can completely change places with material on another line of force.

A motion of this sort will not require any magnetic work to be done, and we must therefore inquire how much of it occurs. Now that we do not have any intrinsic stability of the lines of force attached to the conducting earth, what is the degree of stability, and what are possible sources of instability?

We are certainly forced to think that the greater part of the region is stable, owing to the particle fluxes and hot gases enmeshed in that field. It is clear both from magnetic measurements and from the Van Allen flux that the system is magnetohydrostatically stable.

While a magnetic storm is taking place the magnetohydrostatic stability of the system may well be destroyed by the effects of that phenomenon. Therefore during a magnetic storm the material and fluxes that live on the various lines of force may be interchanged in such a way as to replace arcs of force, one by another, over a large interval of height.

Although most of the region is undoubtedly magnetohydrostatically stable most of the time and therefore does not convect, there is a region close to the earth, an inner region at low latitudes, where the diurnal heating of the outer atmosphere by the sun would be likely to produce an extra driving force for convection by supplying heat from below. Just as in a compressible liquid heated from below, this leads to magnetohydrostatic instability and to a convection in which tubes of force and the material on them would replace one another but without changing the strength of the field anywhere.

Figure 6 is a graph reproduced from Van Allen's results [*Van Allen,* 1959]. It shows a minimum in the radiation zone and an apparent separation into two zones, a feature relevant to our argument. The supply of particles to the outer region is most probably from outside but is the inner one due to some other cause?

If we look into a variety of geophysical data including magnetic field effects, the Van Allen radiation, and the night airglow, we find that a region of latitude at 40° is singled out as a special zone of transition. In the Van Allen zone, 40° is the latitude where the lines of force emerge that go through minimum radiation intensity. As a second example [*Chapman and Bartels,* 1940], in Figure 7 we see also that at 40° a reversal occurs in the sign of the diurnal variation in the earth's field.

I would suggest that in fact all the region within this zone, but not outside it, is caused to convect diurnally at a comparatively slow speed, although most rapidly of course in the tropical zone. This convection is driven by e

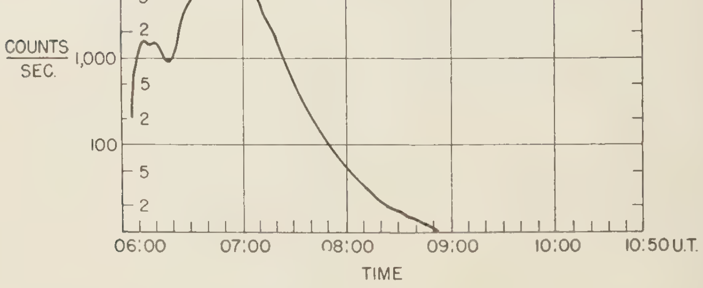


Fig. 6—Intensity measured in the radiation belt during the Pioneer I flight showing the separat into two zones [*Van Allen,* 1959]. The magnetic field line passing through the minimum in the count rate intersects the surface of the earth at geomagnetic latitude 40°.

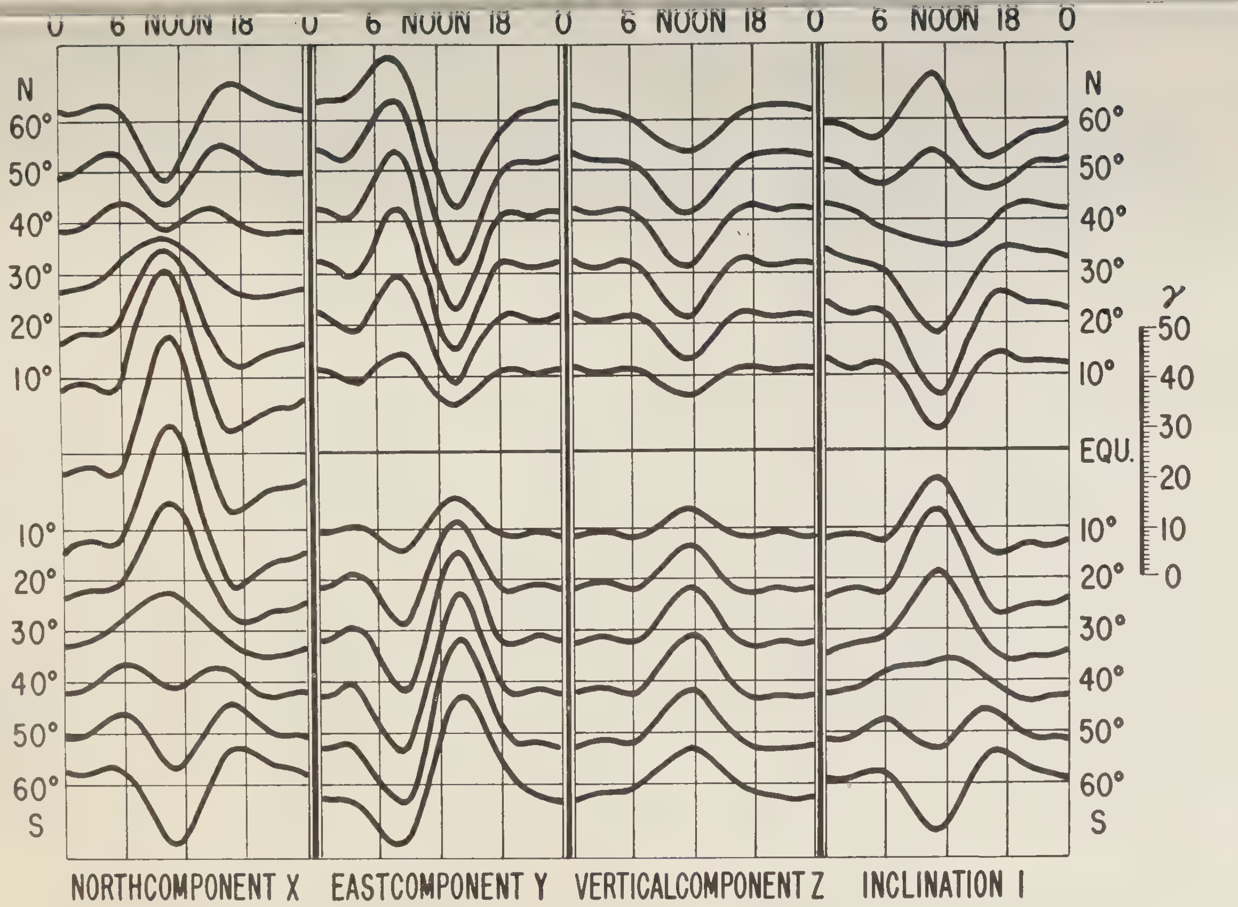


FIG. 7—Diurnal variation in the geomagnetic field, indicating a change in sign at latitude 40° [*Chapman and Bartels*, 1940].

ergy supplied in the form of solar-heated atmospheric gases which constitute a source of heat from below and which produce a magnetohydrostatic instability. The changeover from a convective equilibrium to a stable equilibrium occurs along the boundary defined by the line of force at about 40°.

I would suggest that the whole phenomenon of the inner radiation zone, the increase in hardness of the particles in the inner peak, and the concentration in total energy flux is connected with the fact that the inner region is diurnally convecting.

If so, there are many ways in which further experimentation will be able to show it. I think that in a relatively short time experiments will clearly demonstrate what kind of convective motions are actually occurring in the earth's field, and whether the type occurs within 40° that would be necessary to account for the characteristics of the Van Allen zone.

References

Bailey, D. K., *J. Geophys. Research, 62,* 431, 1957.

Chapman, S., and J. Bartels, *Geomagnetism,* Clarendon Press, Oxford, 2 vols., 1940.

Van Allen, J. A., *Nature, 183,* 430, 1959.

Discussion

Question: Is it possible to estimate the magnitude of the convective velocities that you require for the lines of force?

Mr. Gold: The convective velocity will depend upon the storage time of the particles. But velocities such as are in fact observed as the translational velocities in the ionosphere are already a factor of 10 or more greater than would be needed to mix the flux around. So the observed horizontal velocities that we see in the ionosphere imply a mixing higher up, which is already fast enoug

Chairman: I should like to ask a question. the heating effect produces a motion at a certa point of the line of force, will a similar moti occur along the whole line of force?

Mr. Gold: Yes. Indeed, it is only a complete symmetrical pattern of motion that is permitte one that needs to do no magnetic work. All oth motions would have to do magnetic work. Tl particular one evades it, but requires the comple symmetry of the motion. That, of course, impl that we ought to look now for ionospheric m tions to be symmetrical between the two feet a line of force.

Question: How slowly can you interchange t lines of force from inside and outside and s maintain the equilibrium distribution? What is t longest time you can assume for turning the ion sphere or turning the magnetic field?

Mr. Gold: That is related to the storage ti of the Van Allen flux. In the inner zone, if y are considering an energy range of flux for whi the storage time might be a matter of a mon you must turn it over a few times in a month.

Question: If the convective motion is due solar heating, you would expect a semidiurn effect in the position of the slot, and as far we can make out from the satellite slot this is n observed.

Mr. Gold: It is not correct that you would e pect a diurnal variation in the position of the s from the convection. That may in fact be ve stably defined by the boundary to which the urnal convection goes. This will be a well defin boundary if the material up above it is stab and it will not show diurnal displacements.

The convection need not be very fast. It mig take perhaps a month to complete the convecti cycle, and little will happen in a day.

There is a further point, namely, whether t inner zone displays an annual or a semiannu variation. In particular a semiannual effect wou be expected. A semiannual effect has been o served in the night airglow and can be readi interpreted by this mechanism.

Extension of the Solar Corona into Interplanetary Space

Eugene Parker

University of Chicago
Chicago, Illinois

I shall first write down a number of obsertional facts, and then in another column mpare them with the results of the theoretical ogram we have been carrying out at Chicago er the past few years. The purpose of this mparison will be to suggest experiments, some which are already in progress, and which will rhaps resolve a few of the difficulties in our esent knowledge of the dynamics of the interanetary medium.

I shall start with the sun, since we believe, Gold and others have pointed out, that it is e source of energy for all phenomena obrved in the earth, in particular magnetic orms and auroras, and the source of energy r all the churning that apparently occurs in terplanetary space. Let me list a few basic cts about the sun that seem most pertinent the problem.

On the one hand, there is the solar corona, hose density at a height of 3×10^5 km may 3×10^7 atoms/cm^3. The corona is rather t; active regions in the corona have temperares, as determined by the Doppler widths of nission lines, of the order of 2 to 4 million grees Kelvin. It is hardly necessary to point t that at these temperatures the gas will be mpletely ionized.

The temperature observations which I quote re are determined only near the sun where e solar corona is sufficiently dense that its nission is visible. Very roughly the temperares have been determined out to 3 solar radii. yond that distance the temperature can only inferred indirectly, and the degree of uncerinty becomes very great.

Somewhat farther out from the sun than the rona itself, in the region where comets are served, at the distance of Mercury or farther, ere is some additional information. Biermann s been investigating over some years now the dynamics of comet tails. In observations on the small knots in the tails of the comet he finds that they experience rather extreme outward accelerations from the sun, much too large to be explained by radiation pressure. Hence, he has inferred that there must be corpuscular streams blowing outward from the sun, which sweep the comet tails away and give the observed acceleration. These same corpuscular streams from the sun that are needed to account for the acceleration of comet tails seem also to be necessary and sufficient for producing the observed ionization and excitation of nitrogen and carbon monoxide in the tails.

In the sense of these indirect inferences Biermann observes gas moving outward from the sun. When the sun is not particularly active, he observes velocities of the order of 500 km/sec and densities of the order of 100 particles/cm^3 at the orbit of the earth. When the sun is more active, he observes an increase in the velocity of the streams to perhaps 1500 km/sec, and a density of perhaps 10^4 particles/cm^3. This last estimate is very rough.

As has been pointed out in a previous paper, the gas is ionized and a good conductor of electricity, and therefore conveys with it whatever magnetic fields are present.

The Babcocks, looking at the Zeeman effect in certain photospheric lines, find that the sun has, at least toward the polar regions, a dipole-like field of about 1 gauss. They find that the equatorial magnetic field on the surface of the sun is also of the order of 1 gauss, but rather disordered. Let me note then that there is a field everywhere over the sun, and usually of about 1 gauss. Near the poles the field looks very much like a dipole. You are all familiar with the pictures of coronal tongues streaming outward from the polar regions, suggesting again that there is a field resembling that of a dipole near the poles and becoming radial at larger distances.

An interesting piece of data has been reported recently by Hewish in observations on the scattering of radio-frequency signals from the Crab nebula. While the Crab is in transit past the sun, Hewish can observe the scattering out to about 20 solar radii, or a tenth of the distance to the earth. As far out as he can observe, he finds that the solar magnetic field is radial. In other words, the radial coronal streamers apparently extend out 20 or more solar radii.

Some further information that can be gained about the magnetic field even farther from the sun than 20 solar radii comes from observations of cosmic rays. In particular, there is the well known event of February 23, 1956, in which the cosmic-ray intensity increased by a factor of 200. The additional cosmic rays clearly came from a flare on the sun. The implications in the manner of arrival of the particles from the sun are very interesting.

The cosmic-ray intensity rose sharply to its peak value at the moment of arrival of the flare particles, and then in a period of an hour began a slow decay. After 20 to 40 hours, depending on the latitude of observation, it had dropped to normal.

The fact that the onset of these particles was so extremely sharp indicates that they must have come directly from the sun. No impeding magnetic field existed to separate their transit times. Therefore one infers that the radial field observed by Hewish at 20 solar radii extended in this case at least to the orbit of the earth.

Examination of the decay part of the cosmic ray indicates that it did not drop off exponentially, but like (time)$^{-3/2}$. This power law is a familiar one. Let us suppose that after the flare the excess cosmic rays were somehow trapped in the solar system, as must have happened, because they continued to arrive for some 30 hours after the flare which produced them had entirely disappeared. Then suppose that the trapping mechanism disintegrates by some process of diffusion. A simple solution is obtained to the diffusion equation in which intensity drops off as $t^{-3/2}$ if the cosmic-ray particles are released in the center of a large three-dimensional space filled with a medium that impedes their flow.

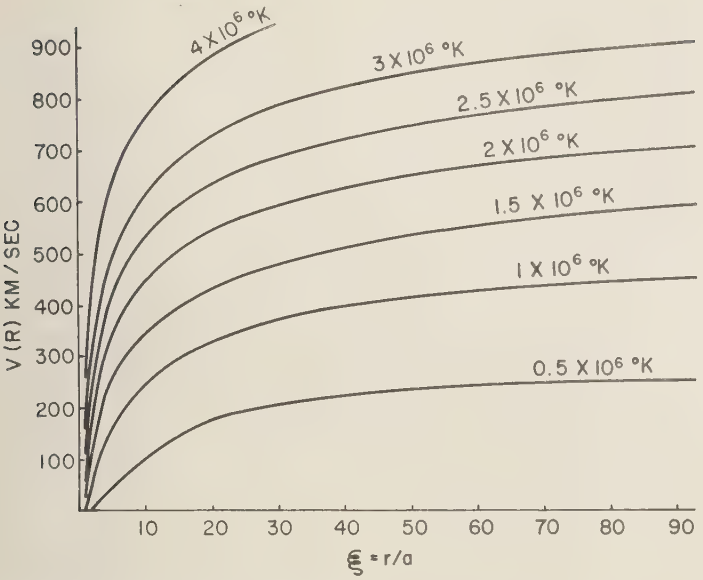
Specifically, the cosmic rays were released at the sun and came quickly to the earth. If outside the orbit of the earth there was a very thick diffusing shell, the intensity at the earth would drop off as $t^{-3/2}$. But no shell is infinitely thick, and when the particles finally reach the outer surface of the shell it can be shown that the solution to the diffusion equation turns down into an exponential decrease. From the point where the cosmic-ray intensity was observed to turn down into a steeper decrease, it may be determined that the shell extended out to the orbit of Jupiter in the event of February 23, 1956. The inner surface of the shell was placed at roughly the orbit of the earth or Mars. The thickness of the shell was therefore 5 astronomical units. If we assume a field strength of 10^{-5} gauss, and a scale of 10^6 km for the disorganized regions in the field, we arrive at the correct rate of diffusion.

This is the history of one of the great cosmic-ray solar flares. As Gold mentioned, there are other flares of more modest type. They do not produce 10-bev particles, but rather a large number of protons of 10 Mev. The particles from these flares also come unimpeded from the sun, again suggesting a radial magnetic field. Again they seem to be stored for fairly long periods of time, as though by an imperfect barrier.

These are the observational facts that I consider most relevant to the problem of the interplanetary gas. Now let me present a brief summary of theoretical results. I shall begin with the solar corona, which is the prime mover for the interplanetary medium.

It has been pointed out that the temperatures in the solar corona are extremely high. When we write down the barometric equations for the corona, we find that the pressure does not vanish at infinity, suggesting that the solar corona must expand continually into space. Therefore we add the term $v\ dv/dr$ to obtain the hydrodynamic equation of motion for the corona. The equation is nonlinear, but it may be solved on the assumption of spherical symmetry in the corona. It is not spherically symmetric; let me make that clear. In fact, there are irregular regions of activity in the corona, but the mathematics cannot be carried through if these irregularities are included. I shall also assume that the solar corona is isothermal radially.

On these assumptions, I have calculated the



G. 1—Spherically symmetric hydrodynamic expansion velocity $v(r)$ of an isothermal solar corona, plotted as a function of r/a, where a is the radius of the corona, taken to be 10^{11} cm.

city of the outward expansion of the solar na, and the resulting gas density at the t of the earth. Results of the calculations wn in Figure 1 indicate that, if the solar na has a temperature of 2.5×10^6 °K to stance of ten solar radii, then the velocity of outward expansion will be slightly in excess 00 km/sec *and* the density at the orbit of earth will be a few hundred ions per cubic imeter, as observed. This velocity agrees with the outward expansion of the solar na observed by Biermann. I suggest there- that the corpuscular streams of Biermann merely hydrodynamic expansion of the r corona. In view of the simple hydrodyic origin of the expansion, it seems appro- te to term the stream a solar wind.

nce the gas of the solar wind is ionized, it carry with it the lines of force of the gen- solar magnetic field. Thus it would seem to account for the radial fields inferred the observed propagation of solar cosmic in interplanetary space, and inferred also the radio observations of the Crab nebula ied out by Hewish.

he sun rotates, of course, so that the field is entirely radial. Figure 2 sketches the degree of spiraling of the field for a solar wind with a velocity of 1000 km/sec.

The field at the orbit of the earth is very nearly radial. At the orbit of Jupiter, the opposite is true. The field is rather tightly spiraled. The field is unstable, because of the anisotropic expansion of the gas as it comes out from the sun. The collision rate is extremely low; hence if the gas expands in directions perpendicular to the radius, but not in the radial direction, its thermal motions become anisotropic and the gas itself becomes unstable. The field becomes disordered. Inserting reasonable parameter values, one finds as a rough result that the field should become disordered between the orbit of Venus and the orbit of Mars.

The field strength, extrapolated outward radially from the sun, and starting with 1 gauss at the sun, gives 2×10^{-5} gauss at the orbit of earth. The characteristics of the instability being known, it may be predicted that the scale of the disordering should be 10^6 km.

We then look back at our earlier inference on the size of the disordered regions in the field deduced from the cosmic-ray data, and we find that there is agreement between our earlier results and the present calculations on the solar

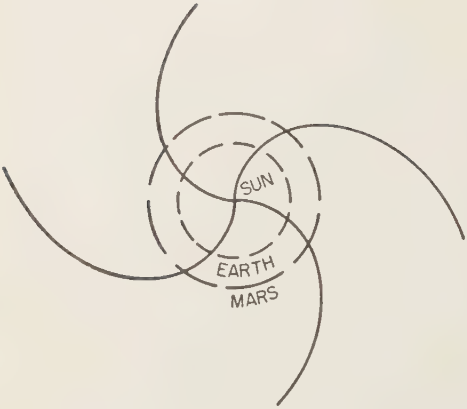


FIG. 2—Projection onto the solar equatorial plane of the lines of force of any solar field which is carried to infinity by outward-streaming gas with a velocity of 10^3 km/sec.

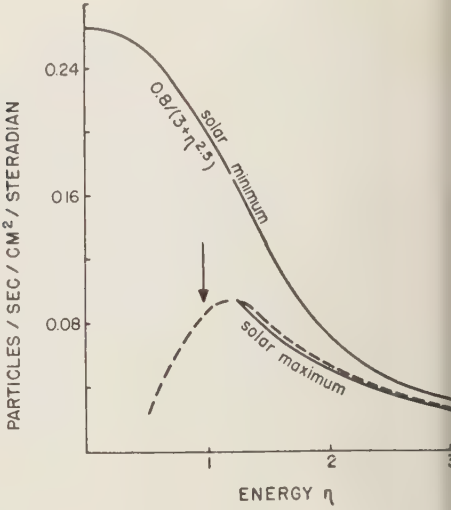


FIG. 3—The solid curves represent the observe primary cosmic-ray differential energy spectru during the 1954 period of minimum solar activi and during the years of maximum activity; tl curve $0.8/(3+\eta^{2 \cdot 5})$ below 1.2 bev is a qualitativ representation of the low-energy observations; tl vertical arrow indicates the energy at which tl spectrum was observed to drop off rapidly durir solar activity. The broken line is the spectru which results from the solar wind in the heli centric shell of disordered field.

wind. Apparently we live inside the thick shell of a tenuous disordered field, extending from the orbit of the earth to the orbit of Jupiter. The fact that the disordered field is being convected outward by the solar wind, at 500 to 1500 km/sec, should have some effect on the cosmic-ray intensity. We believe that most of the cosmic rays come to us from interstellar space, and to get to the earth they must come through the disordered shell. Yet, if they are continually being swept outward by that shell, their intensity should be decreased inside the shell, that is, in the vicinity of the earth.

By assuming that the particles walk randomly through the disordered field, we obtain a diffusion equation that is easily solved. Figure 3 gives the results of the calculations for a solar wind of 1000 km/sec. The upper curve is the differential cosmic-ray spectrum at solar minimum, when the sun is relatively inactive, and we probably see a spectrum resembling that of the true galactic cosmic rays. The lower solid curve is the observed cosmic-ray spectrum during the years of peak solar activity. We see that the intensity, particularly at low energies, is very much depressed.

The vertical arrow indicates where the cosmic-ray intensity seems to cut off very sharply. The broken line represents the theoretical depression of the upper curve, the galactic cosmic-ray energy spectrum, by a 1000 km/sec solar wind in the disordered field beyond the orb of the earth. It resembles the spectrum observe during times of solar activity. Hence this ma be the origin of the 11-year cycle in the cosmi ray intensity.

One may ask how these theoretical proposa can be tested. I think the most crucial pie of evidence will come not too long from no through observations that are being planned b Professor Rossi at MIT. Rossi proposes measure directly the gas blowing outward fro the sun, by a plasma probe to be flown in a veh cle included in the NASA program. The app ratus will determine the density and energ spectrum of both ions and electrons by sophisticated form of ion trap. The observatio will tell us in detail the solar wind velocit density, temperature, and time variations. W should be able to proceed with a more elabora theoretical model than the spherically syn metric one already treated.

A second important experiment is the accu- e determination of the cosmic-ray variation. e cosmic rays vary with the 11-year-cycle. ave suggested very tentatively that this may a heliocentric phenomenon, that is, that the iation probably occurs over the entire inner ır system. However, other events, such as Forbush decrease, appear to be local phe- nena. For example, the Forbush decrease ws geographical irregularities, and therefore east a part of it must be geocentric. Webber, Maryland University, who has been studying observational data, suggests that the For- h decrease may be a combination of helio- tric and geocentric effects. It would be very eresting to have detectors spaced near the th and at various distances out from the th to determine the precise separation of mic-ray variations into geocentric and helio- tric phenomena.

n the remainder of this paper I shall discuss terrestrial effects of the solar wind.

should like to note first that a solar wind wing outward from the sun must exert a ssure against the geomagnetic field, inas- ch as the ionized gas of the solar wind is ble to penetrate the field.

ext, I should like to note the instability of interface between the geomagnetic field and solar wind. Suppose that the magnetic field onfined to the region below a certain plane, ve which there is a beam of ions represent- the solar wind. If the plane surface is then turbed it can be shown that waves will ap- r in the surface and grow with time, dis- ying the analogue of the classical Helmholtz ability. A detailed investigation of the in- ility indicates that the waves in the surface the geomagnetic field will have a velocity ıparable to the solar wind velocity. The elength will be only slightly larger than Larmor radius of the particles in the geo- netic field, i.e., a few hundred kilometers.

f the outer surface of the geomagnetic field is table, it follows that the region of the inter- will be disordered. One may then ask, to ıt extent will the gas from the solar wind rder the field and push aside the lines of gnetic force, and how deeply will the solar d insert tongues of gas between the lines of e of the geomagnetic field?

A rough estimate may be obtained by supposing that the tongue can probably penetrate to that depth in the geomagnetic field where geomagnetic pressure is equal to the wind pressure. With this assumption one finds that the normal solar wind, which has a velocity of 500 km/sec at a density of 100 particles/cm^3, can penetrate no deeper than to within 5 earth's radii from the center of the field. Outside 5 earth's radii the field must be strongly agitated. The enhanced solar wind of 1500 km/sec and 10^4 particles/cm^3 can penetrate to within perhaps 2 earth's radii.

Now consider what might happen to an individual proton in the region of disordered fluc-

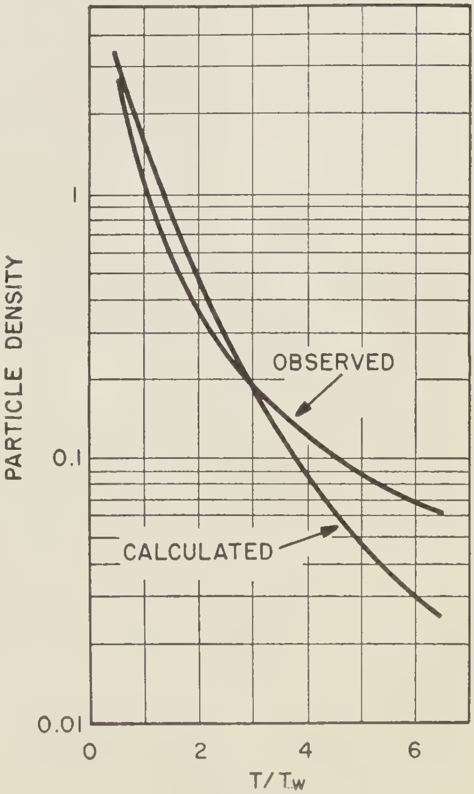


Fig. 4—Comparison of the particle spectrum derived from the Fermi acceleration mechanism (calculated) with the observed primary auroral proton spectrum (observed). T_w is the kinetic energy of an ion moving with the velocity of the solar wind, in this case taken as 2000 km/sec.

tuations of the penetrating tongues. We expect that the Fermi mechanism will be operating. Since we cannot write down the field in detail, we cannot give a rigorous result for what will actually happen to the particle. But we can put a lower limit on the acceleration that will take place by representing the transition region between the geomagnetic field and the solar wind as a simple smooth shear in which ropes of magnetic flux are embedded. The calculated curve in Figure 4 shows the results of a calculation based on the above model. It is derived from a Fokker-Planck equation wherein the particles are random-walked out of the region while they are simultaneously undergoing Fermi acceleration. The curve in Figure 4 is based on a high solar wind velocity of 2000 km/sec.

The observed curve in Figure 4 represents the spectrum of primary auroral spectrum for protons as deduced by Chamberlain from observations of the aurora. The agreement between the two curves is reasonably good and suggests that the proposed mechanism may indeed be the origin of the auroral particles.

The Fermi mechanism does not appear to accelerate electrons by interesting amounts. But there is another mechanism, involving a plasma interaction, wherein the gas from the solar wind, mixing with the gas carried by the geomagnetic field, sets up strong velocity fluctuations in which the electrons receive about half of the ion energy. In this way the kinetic energy of the ions in the solar wind may be converted into the energy of individual electrons, suggesting as an order of magnitude the generation of 20-kv electrons in the interface between the solar wind and the geomagnetic field.

It is interesting to note that the energetic electrons and protons generated in the interface between the geomagnetic field and the solar wind will move freely along the geomagnetic field lines. Since the interface is located at 5 earth's radii, and the field lines passing through the equator at that distance intersect the surface of the earth at a latitude of 65°, we may expect that these particles produced by Fermi acceleration will be concentrated in a belt at 65°, which is in fact the center of the auroral zone.

There seems, then, to be a simple connection between the auroral latitudes and the depth to which the solar wind can penetrate in th geomagnetic field. The Fermi acceleration o particles occurring in the disordered interfac between the magnetic field and the solar win may account for the primary auroral propertie and plasma interactions may perhaps accou for the electrons that participate in auroral a tivity.

It will be noticed that tongues of magnet field from the sun have not been mentione although they play an important role in th consideration of Professor Gold. I believe tha these tongues will not accomplish the objectiv Gold has set out for them. He thinks opposite We will spare you the debate, because an e periment may be proposed to settle the issu If we were to set up an observing station b tween the earth and the sun, at least as dista as the moon, we could detect the particles transit from the sun to the earth, study t structure of the shock waves at the fronts the expanding gas clouds, and perhaps measu magnetic field strengths.

But in particular, with regard to the questi of the magnetic tongues, the particles shou be detected at the outer station before th arrive at the earth, and, if the cloud is movi at 1000 km/sec and the station is at the or of the moon, they will be seen 4 minutes befo they reach the earth. On the other hand, if, as have suggested, there are no magnetic tongu and particles simply come from the sun whe ever they are on the right lines of force—r membering the solar field to be radial—then t delay time will not be 4 minutes, but rath a time appropriate to the actual velocities of t particles. For example, 10-Mev particles mo at a speed of 50,000 km/sec and will trave the moon-earth distance in 8 seconds.

A second experiment of value would be c concerned with the Forbush decrease. I ha previously suggested that measurements of t cosmic-ray intensity at several distances fr the earth would permit a separation of g centric and heliocentric modulating factors. T value of these observations would be grea enhanced if they could be supported with sim taneous magnetic field measurements at ea space station, since most theories of the crease agree in attributing them to magne fields, but the configuration and mode of ope tion of the fields are open questions. In general it is very important to support particle measurements with magnetic field observations.

Discussion

Question: One of the fascinating problems introduced by both of the previous speakers concerned the question of whether the motions of the solar wind or corpuscular radiation are indeed radial from the sun or whether they are spiraling in shape because of the rotation. Comets can give the answer to this question.

The second point is a suggestion in the placing of observing stations. There may be a use for the Lagrangian point in orbits between the earth and the sun, about 1.4 million miles from the earth.

Question: There is a second mechanism for filling the Van Allen layer, which also depends on the solar wind. This is a simple scattering of the solar-wind protons in the outer atmosphere. When the scattering is through an angle of 10° or so, the particles are injected into trapped orbits with high efficiency.

Of course, the soft particles that have been injected in the outer layer are probably electrons. There may be 10-kv protons, however, because these could not be detected with the instrumentation flown thus far.

In the course of time a circulating belt of trapped 10-kv protons would accelerate the electrons in the outer atmosphere to comparable energies. We estimate the time required for such acceleration to be about 10 hours, which is reasonable.

So we believe that this very simple scattering mechanism, together with the hypothesis of the solar wind, can provide an adequate injection rate into the outer layer.

With respect to the inner layer, it is difficult to understand why, in times of unusual solar activity, the oncoming cloud does not perturb the earth's field sufficiently to feed soft particles directly into the inner layer. The population of the inner layer will then decay until replenished by another burst of solar activity without calling on the convection mechanism.

Mr. Gold: I do not understand why that would make a maximum. I could understand why the process would bring them down to the inner zone, but not why it would produce a separation between the zones.

I should like to ask Professor Parker what shape he does think would result in the magnetic field if a magnetized region, such as we know to be responsible for a solar outburst in the chromosphere—not in the corona—explodes, in addition to the configuration that he has been discussing in connection with the coronal outstream?

What will be the superposition of fields if we take a magnetic region in the close vicinity of the sun, in the chromosphere, where an outburst occurs such as we see, and material is flung out?

Mr. Parker: In the first place, the configuration will depend on the circumstances under which that explosion occurs. But to answer your question, let me construct what I think you have in mind. Consider the surface of the sun, and suppose that there is a field associated with a sunspot or some other active region, a fairly strong field, in excess of 1 gauss.

Suppose that underneath this there is an explosion which blows particles into space. I think this is what you have in mind. In that case, after the explosion, the lines of force will have just the configuration that you indicated in your drawings.

The question is now: will this configuration accomplish the objectives you proposed?

Mr. Gold: This will merely replace the other field that was previously there, will it not?

Mr. Parker: It will displace the radial field.

The Geomagnetically Trapped Corpuscular Radiation

JAMES A. VAN ALLEN

State University of Iowa
Iowa City, Iowa

ntroduction—One of the most interesting physical discoveries of recent years was that de with the early United States satellites plorer I (satellite 1958$_\alpha$) and Explorer III tellite 1958$_\gamma$). It was found [*Van Allen,* 8; *Van Allen and others,* 1958] that an imnse region around the earth is occupied by ery high intensity of charged particles (pros and electrons), temporarily trapped in geomagnetic field. The detailed study of radiation has been a major endeavor of the t year and a half by a group at the State versity of Iowa in the United States, and IGY workers in the Soviet Union. Important itional information at relatively low altites has been obtained in rocket experiments rn by other workers. Although knowledge of trapped radiation is still incomplete, very stantial progress has been made in observing interpreting this new phenomenon [*Coleman others,* 1959; *Gold,* 1959; *Kellogg,* in press; *en and others,* 1959; *Singer,* 1958*a* and *b*; *Allen and Frank,* 1959*a*; *Van Allen and ers,* 1959; *Vernov,* 1958; *Vernov and others,* 9*a* and *b*].

he reader is referred to *Van Allen and others* 59], *Van Allen and Frank* [1959*a*], and *Ver- and others* [1959*b*] for the most complete ounts available in print.

nderstanding of the dynamics of the trapg of charged particles in the geomagnetic has also been considerably advanced by the us experiments of August–September 1958. se experiments comprised the artificial inion of beta-decay electrons from the fission ments produced by small nuclear explosions igh altitudes and the subsequent observation geophysical effects by Explorer IV (satel- 1958$_\epsilon$), by sounding rockets, and by other niques of the geophysical effects produced *tional Academy of Sciences,* 1959].

Figure 1 gives an over-all view of the intensity structure of the two principal radiation zones (or belts) around the earth as derived from our Geiger-tube observations [*Van Allen and others,* 1959; *Van Allen and Frank,* 1959*a*] with Explorer IV and with Pioneer III. Generally confirmatory results have been obtained by the Soviets with radiation instrumentation on Sputnik III and Mechta [*Vernov,* 1958; *Vernov and others,* 1959*a* and *b*].

An immense amount of more detailed information at the lower altitudes is available in the full gamut of Explorer IV observations with a variety of detectors [*Van Allen and others,* 1959]. These observations are still under intensive study.

The most recent data [*Van Allen and Frank,* 1959*b*] from Pioneer IV on March 3, 4, 5, and 6, 1959, again confirm the general structure of the region of trapped radiation but show a very great enhancement of intensity and a considerable expansion of the outer zone following the great M-region event on the sun on February 25, 1959 (Fig. 2). In addition these data provide information on the absorptivity of the radiation in both inner and outer zones and give a new determination of the cosmic-ray intensity in interplanetary space.

Nature of the trapped radiation—On the basis of the extremely rapid increase of intensity with altitude above some 1000 km, it was immediately evident [*Van Allen,* 1958] that the newly discovered radiation must consist of charged particles constrained by the earth's magnetic field, since the additional atmospheric absorption between altitudes of, say, 1000 and 2000 km was several orders of magnitude less than the wall thickness of the detectors used in 1958$_\alpha$ and 1958$_\gamma$. There is now a wealth of evidence supporting this conclusion in full detail, and we regard it as established beyond

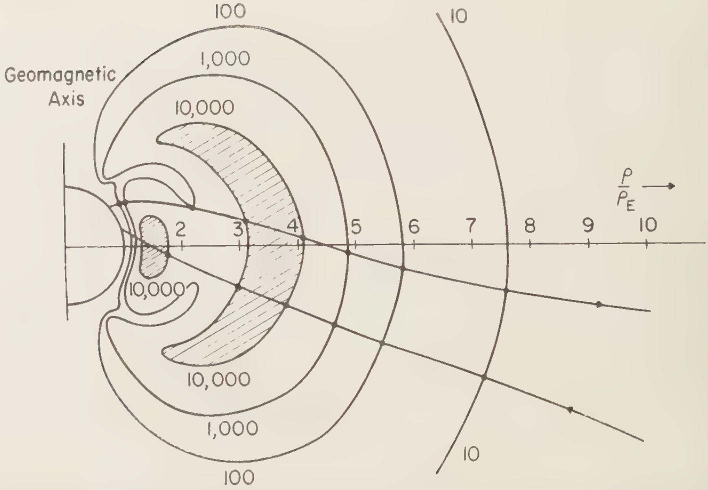


FIG. 1—Intensity structure of the trapped radiation around the earth. The diagram is a geomagneti meridian section of a three-dimensional figure of revolution around the geomagnetic axis. Contours c constant intensity are labeled with numbers 10, 100, 1000, and 10,000. These numbers are the true count ing rates of an Anton 302 Geiger tube carried by Explorer IV and Pioneer III. The linear scale of th diagram is relative to the radius of the earth, 6371 km. The outbound and inbound legs of the traje tory of Pioneer III are shown by the slanting, undulating lines. The intensity structure is a function detector characteristics. (See text.)

all reasonable doubt that the observed radiation consists of charged particles trapped in the earth's magnetic field in the manner visualized by Poincaré, Störmer, and Alfvén in classical theoretical studies.

The nature of the particles and their detailed energy spectra are much more difficult to establish conclusively. We originally reasoned as follows:

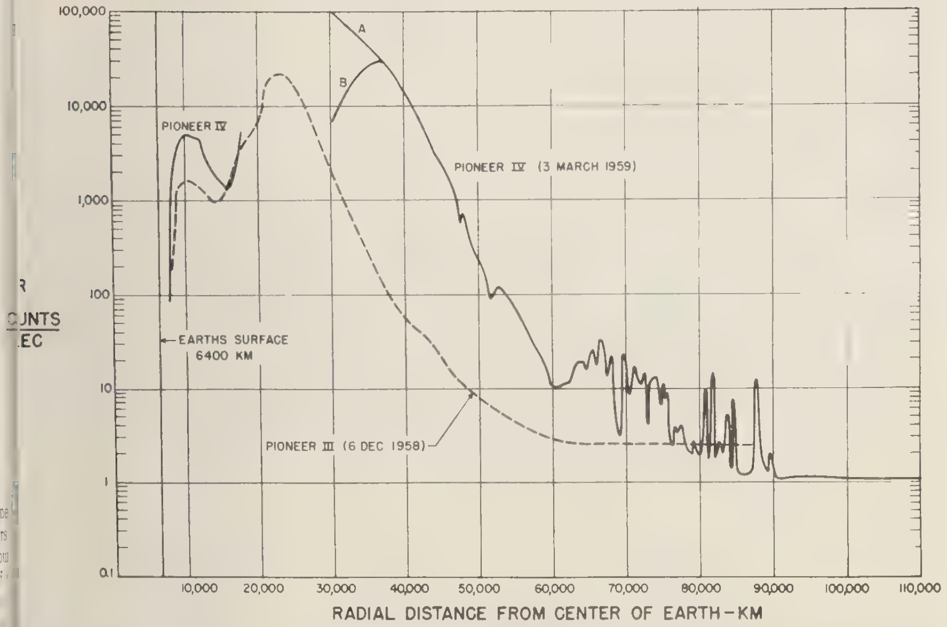
On the grounds of their universality in nature it is reasonable to suppose that electrons are present. Among the many nuclear possibilities, protons are likely dominant because of the preponderance of hydrogen as an atomic constituent of matter. The simplest working hypothesis, then, is that the trapped radiation consists of electrons and protons. If this assumption is granted, the problem becomes one of measuring the absolute differential energy spectrum of each of the two components as a function of position in space, direction, and time.

It is immediately evident from the extensive Explorer IV observations [*Van Allen and others*, 1959] that there is no simple, universal answer to the above problem. The intensity of the radiation *and* its composition vary strongl with position in space, direction, and time. thoroughly satisfactory study of the proble must await more elaborate experimental obse vations.

Meanwhile, the available observations, ol tained with equipment relatively rudimenta by laboratory standards, have served to pr vide a good reconnaissance of the nature of tl trapped radiation. Indeed, it is my opinion th the present state of knowledge is such that tl nature of the radiation, when finally determine will prove to be of interest and value but prol ably will not be markedly different from tl following sketch:

(*a*) The observations with the diversity detectors carried by Explorer IV, Sputnik II Pioneer IV, and Mechta demonstrate concl sively that the nature of the radiation (its cor position and the energy spectra of its cor ponents) in the inner zone is quite differe from that in the outer zone.

(*b*) The integral range-spectrum of the rac ation in the inner zone falls by two orders magnitude from 1 mg/cm² to about 140 mg/cn



IG. 2—A comparative plot of the radiation intensities as measured with nearly identical Anton 302
ger tubes in Pioneer III and Pioneer IV. The trajectories were not identical, the most important
erence being that Pioneer IV cut through the inner zone several degrees closer to the equator, at a
ial distance of about 10,000 km.

n. then trails out more gradually toward
ater stopping powers. Of the radiation which
etrates 140 mg/cm² a fraction of 1 per cent
penetrates several grams per square centi-
ter. On the basis of crude range and specific
zation measurements in Explorer IV, the
re penetrating component is tentatively iden-
d as consisting of protons having energies
he order of magnitude of 100 Mev. The less
etrating radiation has a low specific ioni-
on and hence quite likely consists of elec-
ns. These electrons may have energies rang-
up to about 1 Mev, and a spectrum rising
ngly (though not as rapidly as that of the
oral soft radiation) toward lower energies.
rgy fluxes as high as 100 ergs/cm²/sec/stera-
n have been found beneath an absorber 1
/cm² in thickness at altitudes of 2000 km
r the geomagnetic equator. As measured with
hin (0.92 g/cm²) CsI crystal [*Van Allen*
others, 1959], more than 0.95 of the energy
: is in the less penetrating electronic com-
ent.

(*c*) The outer zone has a quite different nature. All evidence is consistent with an exclusive electron content, as far as the characteristics of detectors thus far used permit observations. The smallest value of absorber used in these detectors is 1 mg/cm². The energy spectra apparently resembles that of the auroral soft radiation, rising sharply toward low energies from an upper limit of about 100 kev. The omnidirectional flux of electrons of energy greater than 20 kev is of the order of 10^{11} particles/cm²/sec in the heart of the outer zone, on the basis of the simplest interpretation, though not conclusively the only possible one. [*Van Allen and Frank*, 1959*b*; *Vernov and others*, 1959*b*].

(*d*) The inner zone is relatively stable as a function of time during the period of available observations.

(*e*) There are marked temporal fluctuations in the 'slot' between the two zones, and variations of very great magnitude in both intensity and spacial structure in the outer zone. These fluctuations are apparently associated with solar activity [*Van Allen and Frank*, 1959*b*].

(*f*) Several recent rocket experiments into the lower fringe of the inner zone are of considerable interest (L. Allen, Jr., R. B. Walton, W. A. Whitaker, and J. A. Welch, Angular distribution of Van Allen radiation with respect to geomagnetic field, private communication, May 1959; S. C. Freden and R. S. White, Protons in the earth's magnetic field, private communication, May 1959; F. E. Holly and R. G. Johnson, Composition of radiation trapped in the geomagnetic field at altitudes up to 1000 km, private communication, May 1959).

Holly and Johnson, using a simple magnetic spectrometer, find that at 1000 km altitude over the central Atlantic Ocean the particles having a range greater than that of 160-kev electrons are predominantly electrons, possibly 3 to 7 per cent protons being present. Of the electrons in the energy range 30 kev to 4 Mev, 99 per cent have energies less than 600 kev. Freden and White have flown a package of heavily shielded (6 g/cm^2). photographic emulsion to a maximum altitude of 1230 km (also on the southeasterly missile range from Cape Canaveral, Florida). By a standard range-and-specific-ionization-measuring technique, they find only protons under the specified absorber. The spectrum in the range of energy E 75 to 700 Mev is well represented by $E^{-1.8}$. The results summarized in this paragraph are consistent with the earlier observations in Explorer IV and are considerably more decisive with respect to particle identification. Absolute values of intensity and the spatial dependence of intensity are best provided by Explorer IV data.

As was remarked earlier, the relative composition of the radiation is a function of position. For example, the radiation in the upper part of the inner zone is considerably softer than that in the lower part, and the radiation in the outer zone is almost completely absorbed by 4 g/cm^2 of lead.

Origin of the trapped radiation—In view of the extensive body of knowledge about 'solar-terrestrial' relationships [*Chapman and Bartels*, 1940; *Mitra*, 1952], and the earlier discovery of the auroral soft radiation [*Meredith and others*, 1955; *Van Allen*, 1957], we originally suggested [*Van Allen*, 1958] that the trapped corpuscular radiation consisted of ionized solar gas which had been injected into the geomagnetic field, perhaps with acceleration to the observed energies occurring as a local phenomenon there.

Subsequently, various workers have proposed that the trapped radiation may arise, at leas in part, from other processes. H. V. Neher (pr vate communication, April 1958) suggested tha a penetrating component may arise from th delayed radioactive decay of μ mesons emerging from the earth's atmosphere as part of th cosmic-ray 'albedo.' N. Christofilos (privat communication, April 1958), *Vernov* [1958] *Kellogg* [in press], and *Singer* [1958*a* and *b* have drawn attention to the neutron componen of the cosmic-ray albedo as a possible sourc and Singer has emphasized the possibility tha such neutrons of high energy may generate th penetrating component in the inner zone. The suggestions depend upon the fact that neutron produced in cosmic-ray-induced nuclear disin tegrations in the atmosphere will move outwar through the geomagnetic field without deflec tion and occasionally undergo radioactive deca in the outer atmosphere. The decay products a neutron (half-life at rest 11.7 minutes) a an electron, a proton, and a neutrino. Th kinetic energy of the proton is comparable that of its parent neutron, and the electron ha a well known beta-decay spectrum with an upper limit of 782 kev for a neutron at rest. Th neutrino cannot contribute to the observed ge physical phenomena.

The observed spectrum and composition the radiation in the *inner zone* resemble thos expected on the neutron beta decay hypothesi the major component being electrons with a energy spectrum resembling that of neutro decay electrons, and the minor (but penetra ing) component being protons with energies the order of 100 Mev. *Van Allen and othe* [1959] present a further discussion of th spectrum.

Many quantitative considerations (sour strength, trapped lifetimes, equilibrium spectr etc.) remain to be examined before the neutro decay origin of the inner zone can be regarded established. Present knowledge favors it, though the source strength (W. N. Hess, Va Allen belt protons from cosmic-ray neutr leakage, private communication, May 1959) the mechanism may be inadequate, and it m be that the inner zone contains an importa

nixture of particles which have been con-
ed inward from the outer zone (T. Gold,
rgetic particle fluxes in the solar system
near the earth, private communication,
il 1959).

here is very little doubt that the great
er zone and the rich variety of associated
ohysical effects, including auroras, airglow,
ospheric heating, and geomagnetic storms,
directly attributable to solar gas injected
temporarily trapped orbits in the geo-
gnetic field. The mechanism for the accelera-
of the particles therein to the observed
rgies constitutes a major unsettled problem.
two immediately evident possibilities are:
acceleration within the sun by the betatron
ct or other mechanisms, and 'piping' to the
h's vicinity along extended magnetic lines
orce; and (*b*) arrival as low-energy particles
^{4} ev) in a cloud of solar gas, intrusion into
earth's field, trapping, and subsequent accel-
tion by magnetohydrodynamic waves or by
er processes in the local environment of the
th. Radiation monitoring in interplanetary
ce for an extended period of time should
inguish between (*a*) and (*b*), although we
y, of course, find that the true situation is
ombination of them.

ime variations—Conditions in the outer zone
e apparently similar during the Pioneer III
hts on December 6 and 7, 1958, and during
Mechta flight on January 2, 1959. But on
rch 3, 1959, during the flight of Pioneer IV
maximum intensity was much greater than
previously been observed and the zone ex-
ded out some 15,000 km farther. This dif-
ence is the most prominent temporal fluctua-
thus far observed. Indeed, it provides the
st striking evidence for the solar origin of
outer zone.

t the lower altitudes of the observations
h Explorers I, III, and IV a wealth of ob-
vational material is available for study of
poral fluctuations. The inner zone is rela-
ely stable, but intensities in the 'slot' be-
een the two zones and in the lower 'horns'
the outer zone are subject to considerable
tuations in a manner that tends to establish
ir connection with other geophysical phe-
nena.

eophysical role of the trapped radiation—
have suggested that the trapped radiation plays an essential role as an intermediate reservoir of charged particles with the sun acting as a source and the earth's atmosphere as a sink. This reservoir may be responsible for auroras, airglow, and atmospheric heating. In addition, the trapped radiation may be the seat of the electrical ring current long supposed [*Chapman and Bartels*, 1940] to be responsible for the main phase of geomagnetic storms.

These views appear plausible on the basis of energetic considerations, the geometric form of the zones of trapped particles, the estimated leakage rates, the particle intensities, and general considerations of plasma physics. But much more detailed observational and theoretical study will be required before each of the many phenomena can be conclusively examined.

Sample tentative intensities—The following results represent preliminary intensity values in the *heart of the inner zone* at an altitude of 3600 km on the geomagnetic equator:

(*a*) Electrons of energy greater than 20 kev: maximum unidirectional intensity $\sim 2 \times 10^9/\text{cm}^2$ sec ster.

(*b*) Electrons of energy greater than 600 kev: maximum unidirectional intensity $\sim 1 \times 10^7/\text{cm}^2$ sec ster.

(*c*) Protons of energy greater than 40 Mev: omnidirectional intensity $\sim 2 \times 10^4/\text{cm}^2$ sec.

Sample intensity values are as follows in the *heart of the outer zone* at an altitude of about 16,000 km on the geomagnetic equator:

(*a*) Electrons of energy greater than 20 kev: omnidirectional intensity $\sim 1 \times 10^{11}/\text{cm}^2$ sec.

(*b*) Electrons of energy greater than 200 kev: omnidirectional intensity $\lesssim 1 \times 10^8/\text{cm}^2$ sec.

(*c*) Protons of energy greater than 60 Mev: omnidirectional intensity $\lesssim 10^2/\text{cm}^2$ sec.

(*d*) Protons of energy less than 30 Mev: no significant information.

Trapped corpuscular radiation around the moon and around other planets—The trapping of charged particles in the vicinity of astronomical bodies is doubtless a quite general astrophysical phenomenon. Indeed, it has already been proposed as a mechanism essential to the dynamics of the solar corona (P. J. Kellogg and E. P. Ney, A new theory of the solar corona, private communication, March 1959). More immediate observational interest perhaps attaches to the possibility of trapped corpuscular radiation around our moon and around other planets.

The emergence of cosmic-ray-produced neutrons from the atmosphere or solid surface of astronomical bodies is probably a universal phenomenon (at least within our galaxy). Also, all the bodies of the solar system are subjected to the impact of plasma from the sun with greater or lesser intensity in accordance with their distances from the sun. Hence, each of the presumably major sources of trapped radiation is present throughout the solar system.

The principal parameters which enter into a general quantitative discussion of trapping are the magnetic moment of the body, the radius of the body, and the density and radial extent of its atmosphere. The mechanism of trapping is now sufficiently well understood that a reasonably confident assessment can be made for any known set of the above parameters. Generally speaking, the greater the magnetic moment and the less extended the atmosphere, the more favorable are the conditions for a high intensity of trapped radiation. There remains the possibility that a mechanism exists for local acceleration of charged particles which is peculiarly favored around the earth, though this seems unlikely.

Consideration of present knowledge makes it appear likely that the moon is surrounded by little or no trapped radiation, since its magnetic moment is expected to be small. The closest observations to the moon (about 61,000 km) with Pioneer IV showed no discernible radiation belt, but this result has only a limited significance at the indicated miss distance.

Mars and Venus may well have substantial radiation belts.

The observational undertaking for settling these matters is one of the most enthralling before us in contemporary space science.

References

Chapman, S., and J. Bartels, *Geomagnetism,* Clarendon Press, Oxford, 2 vols., 1940.

Coleman, P. J., Jr., C. P. Sonett, and A. Rosen, Ionizing radiation at altitudes of 3500 to 36,000 km: Pioneer I, *Bull. Am. Phys. Soc., 4* (4), 223, 1959.

Gold, T., Origin of the radiation near the earth discovered by means of satellites, *Nature, 183,* 355–358, 1959.

Kellogg, P. J., Possible explanation of the radiation observed by Van Allen at high altitudes in satellites, *Nuovo cimento,* in press.

Meredith, L. H., M. B. Gottlieb, and J. A. Van Allen, Direct detection of soft radiation above 50 kilometers on the auroral zone, *Phys. Rev., 97,* 201–205, 1955.

Mitra, S. K., The upper atmosphere, Asiatic Society Monograph Series, vol. V, 2d edition, Calcutta, 1952.

National Academy of Sciences, April 29, 1959, Washington, D. C., Symposium on scientific effects of artificially introduced radiations at high altitudes, *J. Geophys. Research, 64,* 865–938, August 1959; published simultaneously in *Proc. Natl. Acad. Sci. U. S.*

Rosen, A., C. P. Sonett, and P. J. Coleman, Jr., Ionizing radiation detected by Pioneer II, *Bull. Am. Phys. Soc., 4* (4), 223, 1959.

Singer, S. F., 'Radiation belt' and trapped cosmic ray albedo, *Phys. Rev. Letters, 1,* 171–173, 1958*a*.

Singer, S. F., Trapped albedo theory of the radiation belt, *Phys. Rev. Letters, 1,* 181–183, 1958*b*.

Van Allen, J. A., Direct detection of auroral radiation with rocket equipment, *Proc. Natl. Acad. Sci. U. S., 43,* 57-92, 1957.

Van Allen, J. A., Paper presented at joint meeting of National Academy of Sciences and American Physical Society, May 1, 1958.

Van Allen, J. A., and L. A. Frank, Radiation around the earth to a radial distance of 107,400 kilometers, *Nature, 183,* 430–434, 1959*a*.

Van Allen, J. A., and L. A. Frank, Radiation measurements to 658,000 kilometers with Pioneer IV, to be submitted to *Nature,* 1959*b*.

Van Allen, J. A., G. H. Ludwig, E. C. Ray, and C. E. McIlwain, Observation of high intensity radiation by satellites 1958α and γ, *Jet Propulsion, 28,* 588–592, 1958.

Van Allen, J. A., C. E. McIlwain, and G. H. Ludwig, Radiation observations with satellite 1958ε, *J. Geophys. Research, 64,* 271–286, 1959.

Vernov, S. N., Special lecture, Fifth General Assembly of CSAGI, Moscow, July 30-August 9, 1958.

Vernov, S. N., A. E. Chudakov, E. V. Gorchakov, J. L. Logachev, and P. V. Vakulov, Study of the cosmic-ray soft component by the 3rd Soviet earth satellite, *Planetary and Space Science, 1,* 86–93, 1959*a*.

Vernov, S. N., A. E. Chudakov, P. V. Vakulov, and Yu. I. Logachev, Study of terrestrial corpuscular radiation and cosmic rays during flight of the cosmic rocket, *Doklady Akad. Nauk SSSR, 125,* 304–307, 1959*b*.

Discussion

Question: In the asymptotic counting rate at large distances from the earth, there seems to be a significant difference between the Pioneer III and Pioneer IV results. How should that be interpreted?

Mr. Van Allen: That has been a troublesome point so far. The graphs actually show the raw counting rates, which were 2.25 c/s and 1.09 c/s on Pioneer IV. These counters were matched within 20 per cent as nearly as we know, so that we are now forced to believe that there was a real difference in interplanetary cosmic-ray intensity between the two flights.

Question: What are the present results regarding the ratio of the soft components to the penetrating components in the inner belt, and as a second question, are Kellogg and others satisfied that the supply of electrons from the decay of thermal neutrons seems adequate to account for the intensity of soft electrons in the inner belt?

Mr. Van Allen: If we take a range of 1 g/cm^2 as the criterion for dividing the penetrating from the non-penetrating particles, the ratio of penetrating to non-penetrating fluxes is 1:100.

Question: May I comment on the second question. The data on neutron albedo produce results in line with the observations; i.e., the neutron atmosphere can supply the inner belt. I think Kellogg agrees on that.

Question: I should like to observe that, if you let particles leak out, you must occasionally let them leak in by the same perturbations. As an example, in early 1958 there was an aurora visible in the Azores. It seems to me this means the particles were coming in from the outside.

Mr. Van Allen: We had the Explorer I working during the February 11–12, 1958 event. Mr. Oshida has recently been analyzing the data. Actually the lower boundary of the inner zone was shoved down about 350 km on this occasion. So I certainly do agree with you that the inner zone is not a totally quiescent region. Solar activity affects that region, too.

Round-Table Discussion

Chairman: JOHN A. SIMPSON, *University of Chicago*

Participants: BRUNO ROSSI, ALBERT R. HIBBS, ROBERT JASTROW, FRED L. WHIPPLE, THOMAS GOLD, EUGENE PARKER, NICHOLAS CHRISTOFILOS, JAMES A. VAN ALLEN

Mr. Simpson: When it was suggested that a round-table discussion should be held after each of the sessions on formal papers, it was expected that there would be wide-open gaps in the areas of interest and that there would be some question about the details of the experiments and theory.

The officers have done so exceedingly well, however, that I think our round-table talk can be restricted to a discussion of some major controversies, and the experiments that can be done in the near future to throw a little light on these controversial ideas.

Perhaps I could start by saying that one of the central areas of concern is the nature of plasma in interplanetary space, both in the quiescent state and in the transient state where shocks may exist. Considerable thought has gone into this problem, both with respect to theoretical matters and also with respect to experiments from which we could learn something about the nature of the plasma.

I might start then by asking if Mr. Rossi would care to make some comments on what might be done as an experiment.

Mr. Rossi: Perhaps I can just say briefly what kind of instrumentation we are planning for preliminary information on the density, velocity, and direction of motion of the plasma. We heard today of densities of the order of 1000 particles/cm^3, and velocities as high as 10^8 km/sec, leading to a current of 10^{11} particles/cm^2/sec. That is a sizable current. Of course, these may be extreme conditions, and we will wish to design the equipment so that it is sensitive to much smaller currents. In principle the method is very simple. You might just have an external grid and an inner collecting plate. If you make the grid negative and the plate positive, the electrons are deflected in one way and the protons in the other, and you measure the current.

The difficulty with this very simple scheme is the photoelectric effect, which produces an additional and unwanted current. From what we know of the photoelectric effect, this current may be larger than the plasma current.

The solution that we are considering now would enable us also to measure approximately the energy of the particles. We add a second grid, with a negative potential, so that it will repel the photoelectrons, and a third grid to which is applied a sine-wave or square-wave potential, which will modulate the plasma current without modulating the photoelectric current at the same time.

I must say that there will still be a photoelectric current produced by the grid, which, though small, will certainly not be negligible.

The initial advantage of this device is that you modulate only the plasma current and therefore distinguish between the plasma current and the photoelectric current. Moreover, you can use an a-c amplifier, which is easier to build and which discriminates directly against the photoelectric current.

In addition, by varying the amplitude of the modulating potential, you can obtain an approximate measure of the energy of the particles.

We contemplate using two such probes at the two ends of the rotation of the vehicle, and measuring the two currents separately, to obtain information on the direction of the flux.

There are other difficulties associated with the photoelectric effect on the skin of the satellite or space probe itself.

Mr. Whipple: Did you say you expected the photoelectric effect to overbalance the higher mobility of the electron?

Mr. Rossi: Yes, in outer space where the density of the plasma is small, it certainly would.

Mr. Whipple: I would disagree with that.

Mr. Rossi: Did you assume a large bulk velocity?

Mr. Whipple: No, not a large bulk velocity. That will affect the result.

Mr. Rossi: We considered cold plasma moving with a high bulk velocity. I think the main difference may lie in this.

Mr. Jastrow: I think a fair statement is that the two effects are comparable in the normal state of the interplanetary medium.

Mr. Gold: It is quite clear that here is a fascinating and essential area for experimentation. I should like to carry the discussion a little further and ask about the magnetic fields and their measurement in the interplanetary medium. Would one of you care to comment upon this?

Mr. Jastrow: It is interesting to note that the proposal by Professor Rossi seems well suited for installation in a lunar satellite, which would serve a second purpose as an anchored space probe. We would then be sampling the interplanetary plasma with such a satellite, keeping it at a convenient distance from the earth for telemetry. And we do in fact hope, with Professor Rossi, to install such a device in one of the lunar probes or satellites.

Mr. Hibbs: I believe that within a few years we may be able to handle telemetry of the type necessary for such an experiment as this over a distance of several hundred million miles. You could put the plasma probe into an escape orbit from the earth and follow it around the sun and back again, indefinitely.

Mr. Simpson: May we continue on to the discussion of the field measurements that might be considered. Do you have a comment on that?

Mr. Gold: I should like to stress the importance of making field measurements that are capable of resolving the structure of the interface between a fresh solar outburst and the gas that has been there before.

There are such borders. They are pretty sharp. We know that such boundaries exist and are narrow because of the sharpness of the onset of magnetic storms. They are most likely to be associated with a jump in the magnetic field strength. It would be almost impossible that the field should not change across the front. We should have a magnetometer capable of measuring that jump in the same space probe in which we measure the fluxes of particles which might be contained behind the boundary and not be available in front of it. It is most important to have both devices in one probe. It will be a pity to have a magnetic measurement at one time and a particle measurement at another, rather than having them together at the same point in space and time together.

Voice: In the last 2 years good measurements have been obtained at the surface of the earth on the absorption of cosmic noise in the polar regions. I wonder if any of the speakers would care to comment on this measurement with respect to the space program and the possibility of using frequencies above the ionosphere.

Mr. Parker: It will be invaluable to do radio astronomy below the present ionosphere cutoff. To take a simple example, the solar wind streaming past the earth is supersonic and should produce a shock wave. It is difficult to estimate with accuracy the efficiency of generation of plasma oscillations and ultimately the efficiency of radio emission, but a very rough attempt suggests that when the solar wind is enhanced the radio emission from the shock waves near the earth may produce as much as 10 megawatts of power in the 1-megacycle range.

On the other hand, going a little farther from the earth, if the solar corona is heated by any of the mechanical mechanisms that have been proposed, it must produce shocks and therefore radio emissions. Estimates, again very rough, suggest that, between 1 and 20 megacycles, the solar radio spectrum should become non-thermal and perhaps increase inversely with the frequency.

Mr. Gold: There is also the point that all the ionospheric measurement techniques would come into their own in the space of the solar system at frequencies of the order of half a megacycle, that being the critical frequency for the conventional interplanetary electron density. Measurements of reflection from advancing fronts will be possible and measurements of electron density itself by the difference in the radiation resistance of an antenna, depending on whether it is above or below the critical frequency of the plasma outside. All these techniques are orbidden from the ground because of the iono- phere. A wonderful field is opened for a skill- ul radio engineer to discovery by simple radio echniques much of the physics of the plasma.

Mr. Simpson: I wonder whether Professor an Allen would care to say anything about he possible existence of heavier nuclei in the rapped regions. I think there was one sugges- on by Dr. Morrison about looking for α par- cles as a method of probing the solar origin f the inner layer. There are other, heavier nu- ei that one might also look for in trapped adiation.

Mr. Van Allen: Dr. Morrison suggested to us ast summer that a crucial experiment might e the search for helium nuclei in the trapped ones. If helium is present, and in a ratio to pro- ons comparable to that in the sun's atmosphere, e would have a very strong piece of evidence n favor of solar origin.

On the experimental side we don't know any- hing about either way at the present time. I hink it is going to be quite difficult to learn his, but I am very keenly aware of it as a pos- ibility. I think the experiment should be tried. There will be no problem of intensities, but difficult problem of discrimination.

Mr. Parker: What energies do you have in ind?

Mr. Van Allen: I think we should consider α articles in terms of tens of kilovolts, actually. hat is the thing that makes it almost impos- ble right now. It is certainly very, very diffi- ult. But it would be an interesting question o examine the nature of the penetrating com- onent in the inner zone.

Mr. Jastrow: I agree with the remark about he simultaneous measurement of plasma par- cles and magnetic field, and I should like to ay that the very first NASA attempts that will e made, so far as they are firm, will include oth the plasma probe and the magnetometer the same package.

Mr. Gold: The devices should include not nly the probes for the soft plasma but also de- ctors similar to those that have measured the an Allen flux.

Mr. Van Allen: I have one comment about Dr. Gold's point. We have now in Pioneer III plus IV about 50 hours of observation with counters in interplanetary space. In this chosen period nothing happened that was detectable by these counters. Of course, this may be just a very unhappy quiescent period. That is, of course, possible.

Mr. Gold: Fifty hours is not a long period by any means from this point of view. And the indications that we have of the particle bombardment—true in the high-energy range, but it is likely to be true in the low-energy range even more so—are that it is a well defined phenomenon, either a heavy flux of particles or none.

Mr. Hibbs: I should like to ask whether there is a particular importance in performing experiments out of the plane of the ecliptic. So far most of the probes that are being considered I think are in the plane of the ecliptic for obvious dynamic reasons.

Mr. Simpson: That is a very interesting point because it relates to the solar wind and the degree of non-uniformity in the perturbed plane. Would you like to comment?

Mr. Parker: Yes. It would be very interesting to get out of the plane of the ecliptic. Biermann's comet observations are limited by the number of comets that have come by in the last 20 or 30 years, and very few of them have been far out of the plane of the ecliptic.

Mr. Whipple: I believe that the accurate radio meteor experiment will shed some light on this question. The results of Davies and the British radio experiment going down to the 8th magnitude show that there are many meteors in 60° orbits. With radio meteor observations we might get a much clearer picture of the solar wind at high altitudes.

Mr. Parker: One expects that the solar wind far from the ecliptic is probably less than in the ecliptic. Active regions on the sun tend not to sit over the poles but rather to be around the equatorial regions.

Mr. Gold: Moreover, the quiet corona has been plotted out to 20 radii, and at that distance it is enormously more intense in the sun's equatorial plane than about the pole.

Capabilities for Space Research

Homer E. Newell

National Aeronautics and Space Administration
Washington, D. C.

This paper has as its purpose the presentation of a brief survey of the planning that is going on in the area of space research.

Our capabilities for space research depend upon the national interest in and support of such research, upon the tools that can be brought to bear on the effort, and upon the scientists who will devote themselves to the field. We have already obtained strong national support of research in space in the creation of the National Aeronautics and Space Administration. Powerful tools in the form of large rocket vehicles, special power supplies and communications equipment, and extensive tracking and telemetering networks now exist or are being developed with the technical support of a wide variety of missions in space research and exploration. It remains for the scientific community to make the most of the opportunities that lie before it.

At the close of World War II the United States entered the field of rocket upper air research. Although the program began with the WAC Corporal, developed by the Jet Propulsion Laboratory for the U. S. Army, the major impetus came from the V-2 program. Starting in 1946 and running for half a dozen years, the V-2 effort included the firing of some 60 captured German V-2 rockets. At the same time other rockets were under development for the purpose of upper air research. In due course, the Aerobee, Viking, Cajun, and other sounding rockets were introduced into the program. The technique of using balloon-launched rockets for upper air soundings was also developed. All in all, before the International Geophysical Year over 400 sounding rockets were fired by United States agencies for high altitude and upper atmosphere research. During the IGY another 200 sounding rockets were fired, along with the first satellites and space probes.

With these sounding rockets many measurements were made that were impossible at the ground because of the absorbing and distorting effect of the earth's atmosphere. Observations were made on ultraviolet light and X rays from the sun, stars, and interplanetary space; the earth's ionosphere; the earth's magnetic field; the pressure, temperature, density, composition, and winds of the high atmosphere; the aurora; micrometeorites; and cosmic rays. Among the many important results were the recent discoveries of the Van Allen radiation belt and of the fact that the earth is slightly pear-shaped.

The launching of the first artificial earth satellites aroused great interest in space research and exploration. After much national discussion and extensive hearings and consideration by the Congress, the National Aeronautics and Space Act of 1958 became law on July 28, 1958. Section 102(a) of the Act declares "that it is the policy of the United States that activities in space should be devoted to peaceful purposes for the benefit of all mankind." In support of this policy, the Act created the National Aeronautics and Space Administration. In Section 102(c) the Act lists the objectives of aeronautical and space activities of the United States. In paraphrased form, these are: (1) the expansion of human knowledge of atmospheric and space science; (2) the improvement of aeronautical and space vehicles; (3) the development and operation of space vehicles; (4) the study of the potential benefits to be gained for mankind through space activities; (5) the preservation of the role of the United States as a leader in areonautical and space science and technology and in the application thereof to peaceful activities; (6) the interchange of information between civilian and national defense agencies; (7) cooperation with other nations in areonautical and space activities and in

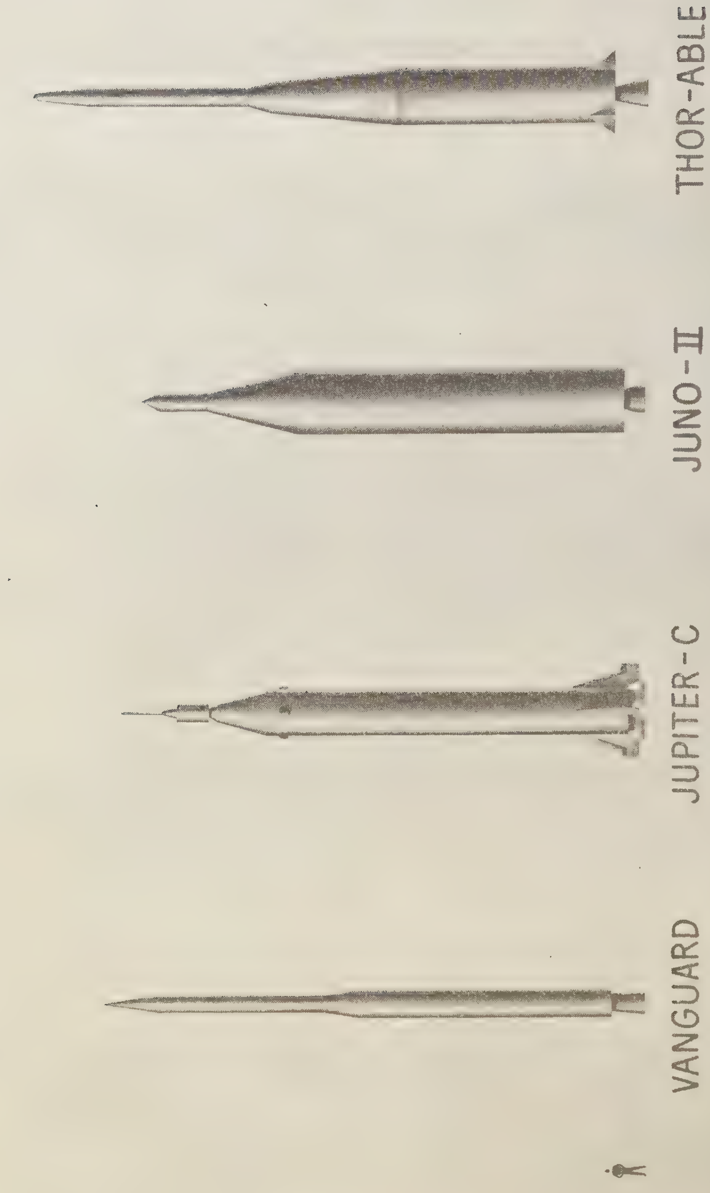


Fig. 1—Existing vehicles in the NASA space sciences program.

peaceful application of the results; and (8) the most effective utilization of the scientific and engineering resources of the United States in achieving these goals.

Under the direction of T. Keith Glennan, the first Administrator, and Hugh L. Dryden, as Deputy Administrator, the National Aeronautics and Space Administration began operations on October 1, 1958. At that time, it took over the personnel and facilities of the 43-year-old National Advisory Committee for Aeronautics. NASA thus began with nearly 8000 scientists, engineers, and technical and administrative personnel and with five field laboratories: Langley Research Center at Langley Field, Virginia; Ames Research Center, Moffett Field, California; Lewis Research Center, Cleveland, Ohio; Wallops Island Station, Wallops Island, Virginia; and the High Speed Flight Station, Edwards Air Force Base, California. There was added to the organizational structure a space flight development department to assume the new responsibilities for the development and operation of space vehicles. Such development and operation will be carried out largely by contract with existing industry and educational groups.

On October 1, the President transferred to NASA from the Department of Defense the original United States scientific earth satellite project, Project Vanguard, with more than 160 scientists and technologists of the Naval Research Laboratory. At that time he also transferred five space probes and three satellite projects that had been under the direction of the Advanced Research Projects Agency of DOD, plus a number of engine development programs from the Air Force and ARPA. Next, on December 3, the President transferred the functions and facilities of the Jet Propulsion Laboratory, Pasadena, California, which had previously been under the Department of the Army.

The NASA program includes a broad area of research and development in aeronautics and space flight. On the space flight side is included the development of large vehicles, of advanced technology, of the applications of space technology to practical problems, and of space science. It is the space science with which this paper is primarily concerned. However, inasmuch as the space science effort depends on the vehicles that are available and on their capabilities, it is worth while to review briefly the national vehicle program. This program has been set up in cooperation with the Department of Defense.

At present, except for the Vanguard vehicle, it is necessary to use vehicles that were designed for other purposes. But with the development of the National Booster Program in the coming years we shall have vehicles that have been designed specifically for space missions.

In Figure 1 are illustrated a number of vehicles that are now available. On the left is the Vanguard vehicle, whose performance capabilities are reflected by the satellites that have been launched—Vanguard I and the meteorological satellite.

Next to it is the Jupiter C vehicle that was used for launching the Explorers. The Juno II, which is based on the Jupiter, is a multistage vehicle using solid-propellant upper stages; it was used in launching the recent Pioneers. The Thor-Able, based on the Air Force Thor and Vanguard upper stages, was used in the early Pioneers.

In Figure 2 are shown vehicles soon to be available. The Scout, on the left, is an all-solid-propellant rocket. The aim of developing this vehicle is to provide an inexpensive means of launching payloads into orbit. The performance capability of the Scout will be something of the order of 150 to 200 pounds in a 300-mile orbit. Because it is built of hardware that either already exists or needs a minimum of additional development, it will be much less expensive than other vehicles, of the order of half a million dollars per shot as compared with two and a half million dollars per shot for the others. The Thor-Hustler is also a Thor-based vehicle, using the Hustler engine for the second stage. The next vehicles, the Atlas-Able and Atlas-Hustler, are based on the intercontinental ballistics missile, the Atlas. I will review their performance capabilities later.

Figure 3 shows the advanced boosters that are now under consideration and on which development work has been started. On the right is the Vega, based again on an Atlas, that is, an ICBM. The second stage will be what is called the Vega vehicle, using liquid oxygen and kerosene as the propellants. The engine for this

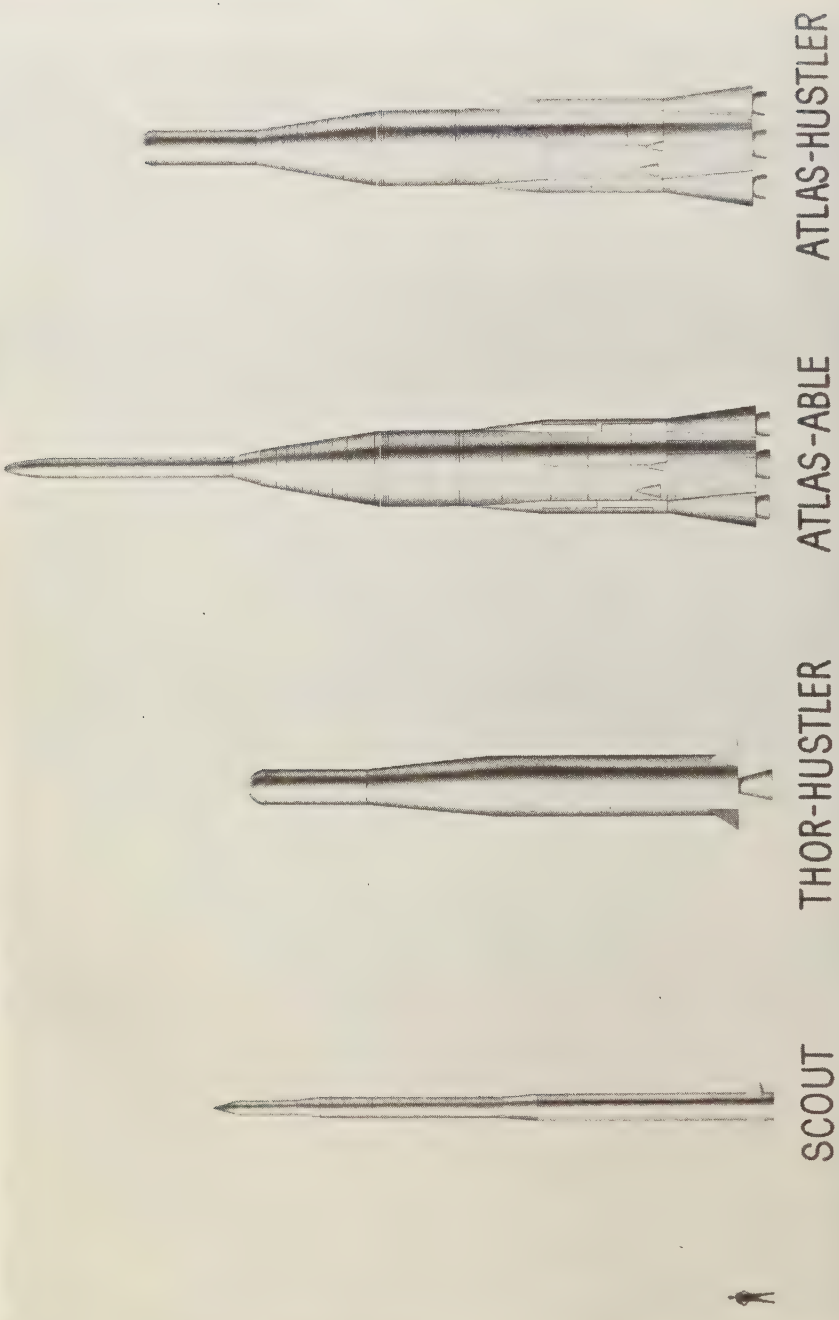


FIG. 2—Vehicles soon to be available for the NASA space sciences program.

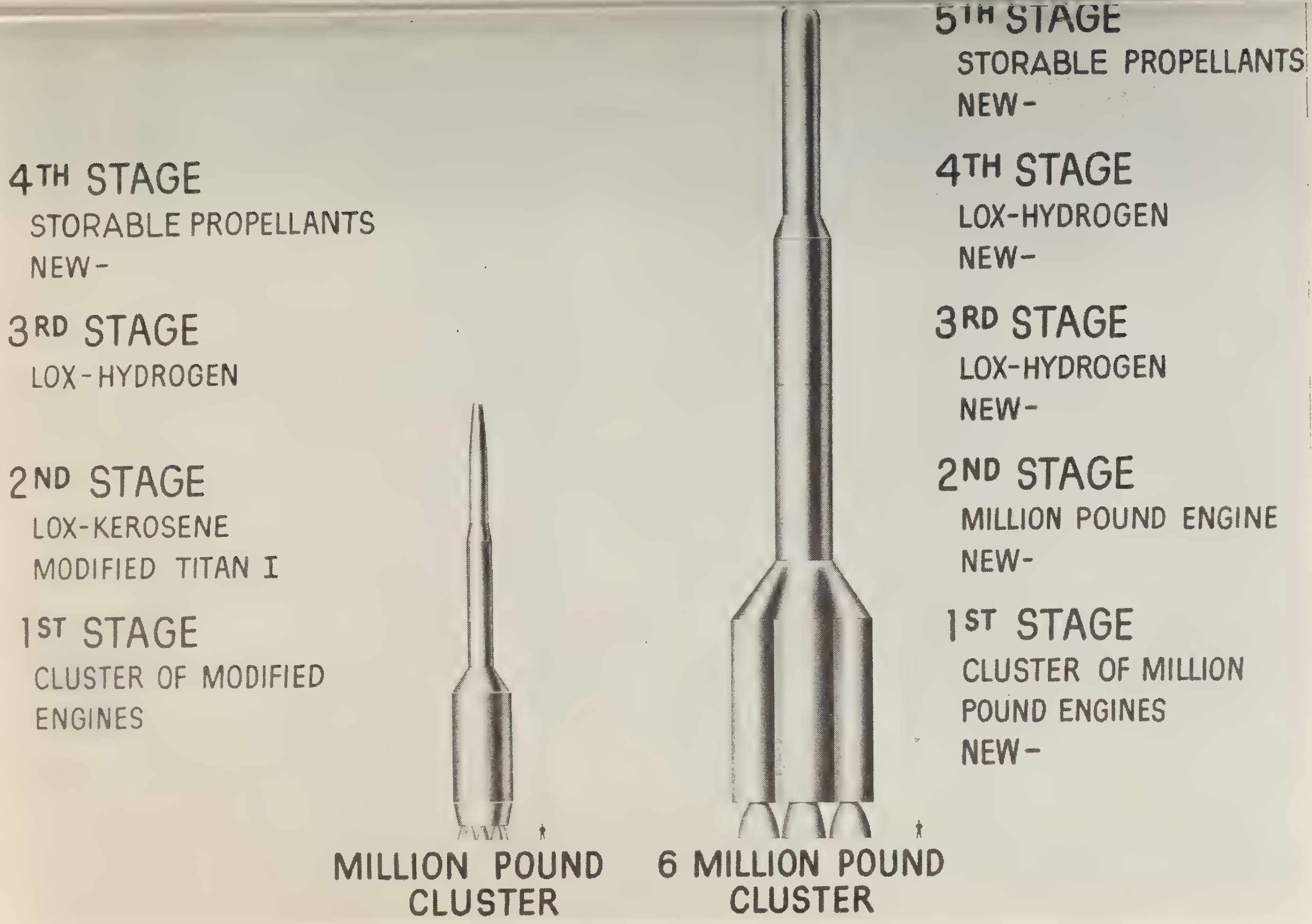


Fig. 3—Advanced boosters in the NASA program.

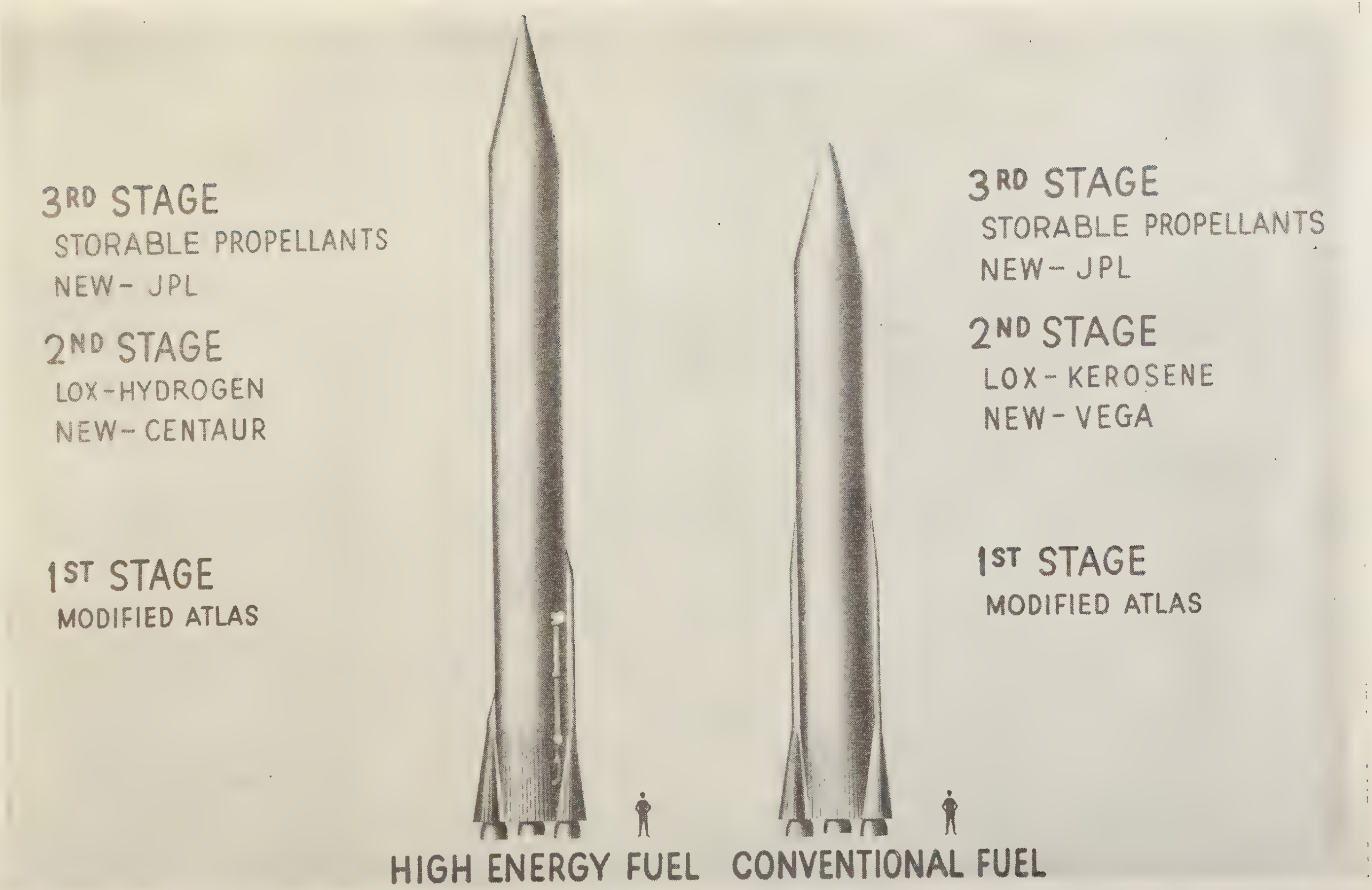


FIG. 4—The second generation of advanced boosters in the NASA vehicle development program. These vehicles will have adequate capability for lunar soft landings and manned lunar mission.

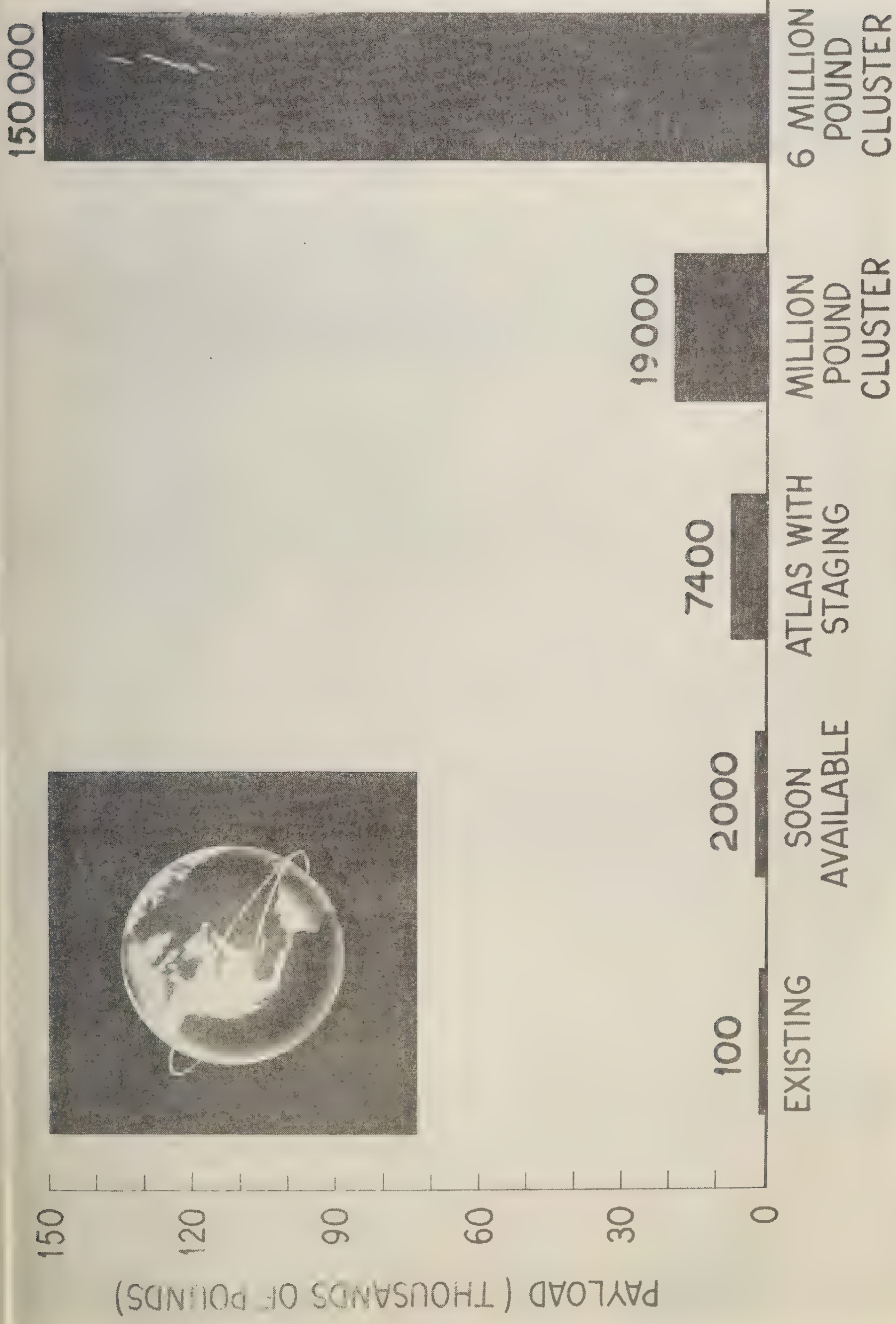


FIG. 5—Payload growth for the 300-mile satellite orbit.

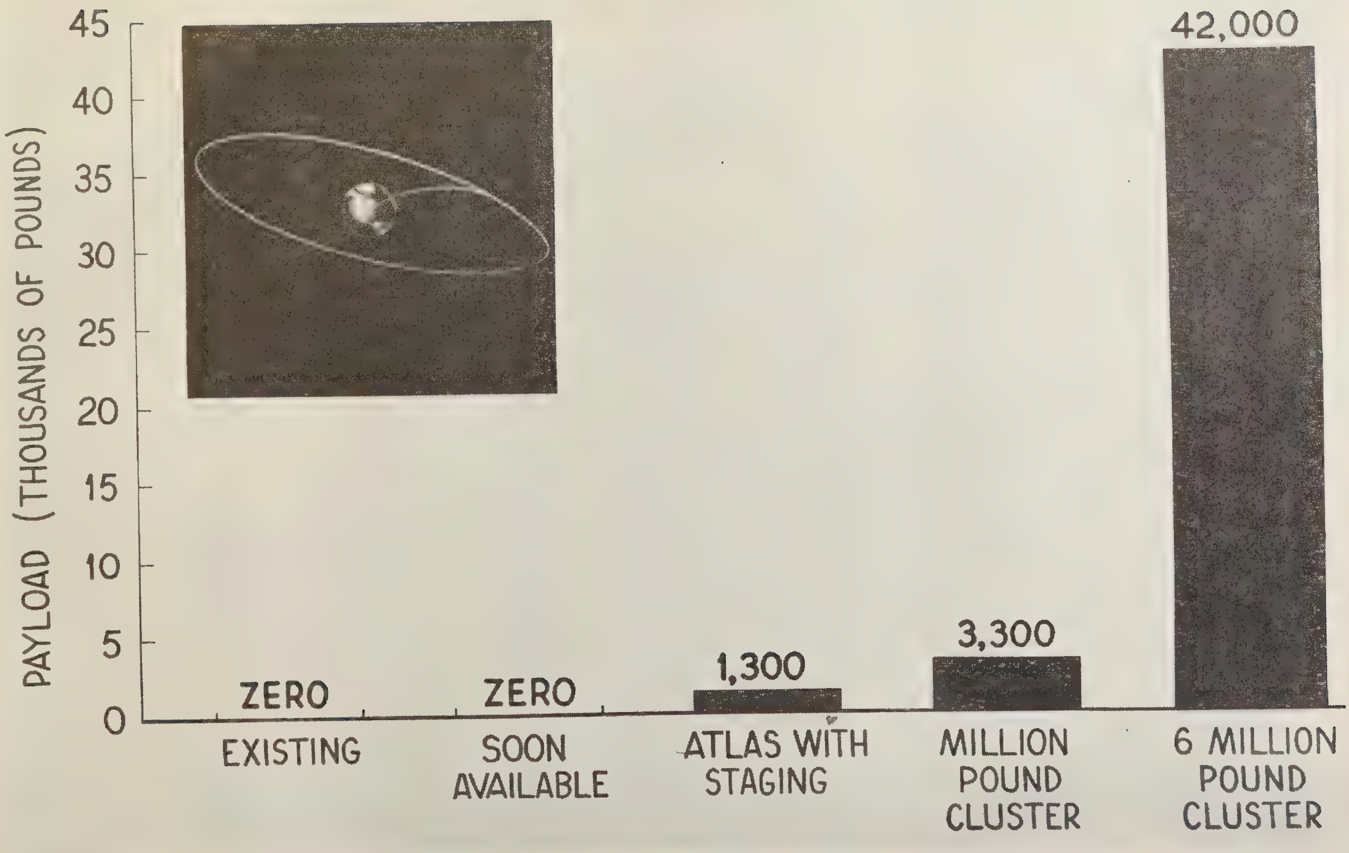


FIG. 6—Payload growth for the 22,000-mile satellite orbit.

is a modified version of the engine that was used for the first stage in Vanguard. The third stage will be a storable-propellant vehicle, developed by the Jet Propulsion Laboratory. The Vega vehicle will be able to put several tons in a 300-mile orbit about the earth. When we have the Vega, a year or a year and a half from now, we shall have the first vehicle that will provide us with a capability equal to that demonstrated by the Soviets in their launching of Sputnik III. Going a step further from the Vega vehicle, we have the Centaur, shown on the left. In concept the Centaur is the same as the Vega vehicle except that the second stage will use high-energy propellants, liquid oxygen and hydrogen.

In the next stage of development is the Saturn, shown on the left in Figure 4. It is a million-pound-thrust engine, clustering, I believe, six to eight existing engines to provide a vehicle that will have multiton capabilities. This vehicle is being developed by the Army Ballistic Missile Agency. On the right is a 6-million-pound-thrust vehicle, based on the clustering of 4½-million-pound-thrust engines. The engine for this vehicle, which is called Nova, is under development; the development lead time is about 4 years. The construction of the vehicle will follow, sometime after the engine becomes available. With the Nova, however, many large-thrust, large-payload requirement missions can be carried out. We can review these capabilities in Figures 5, 6, and 7.

Let us take as our standard a 300-mile orbit as shown in Figure 5. At present we can get small payloads into this orbit. With the vehicles listed as 'soon available,' we can get 2000 pounds into it. When we go on to the Atlas vehicle with staging, the Atlas-Hustler, and the high-energy stages, the Vega and the Centaur, we will get on the order of 7000 pounds in a 300-mile orbit. With the million-pound cluster, the Saturn vehicle, 19,000 pounds will be possible; with the Nova, 150,000 pounds. It is at this time that we shall be able to think of creating manned laboratories of considerable size.

In Figure 6 the performance capabilities are referred to the 22,000-mile orbit. You will recall that this is the orbit at which the period of revolution is exactly 1 day. At present we have no payload capability for this orbit. In the near future, in the vehicles that are on the way now, we shall still not have this capability. But the Atlas with staging is capable of placing 1300 pounds into the 22,000-mile orbit; the million-pound cluster vehicle, the Saturn, can put a ton and a half and the Nova vehicle can put 42,000 pounds into such an orbit.

In Figure 7 we relate the vehicle capabilities to the ability to land a capsule on the moon. At present there is no capability. In the vehicles 'soon available,' the expected capabiilty is about 300 pounds; with the multistaged Atlas, it will be 730 pounds, and with the million-pound cluster, 1800 pounds. This is the weight *landed on* the moon. Not until we have a vehicle in the class of the Nova can enough poundage be landed on the moon to permit return from the moon with appreciable weights.

Nuclear rockets for application to space missions are now being developed jointly by the Atomic Energy Commission and NASA; when available, they will have still greater capabilities, but the time scale is probably even longer than that given for the previous vehicles.

Working with groups already engaged in upper atmosphere and space research during the International Geophysical Year, and with others who have expressed interest in space research, the NASA is developing a space science program that picks up and increases the momentum acquired during IGY. For convenience we have divided the program into a number of areas, and I should like to show very quickly a number of graphs indicating those areas.

We first consider the study of atmospheres. Figure 8 suggests some of the complication of our own earth's atmosphere. Above the conventional weather region in the upper atmosphere we have weather conditions, if they may be referred to as such, which are based mainly on chemical and electrical changes. Shown in the figure is the sun, which is, of course, the primary source of energy driving the atmosphere. Our interest, however, is not just in the earth's atmosphere; we are interested also in the atmosphere of the sun, the moon, and the planets—hence the 's' on 'atmospheres.'

Figure 9 shows the ionosphere. Note the difference in complication between the daytime ionosphere shown on the right and the nighttime ionosphere shown schematically on the

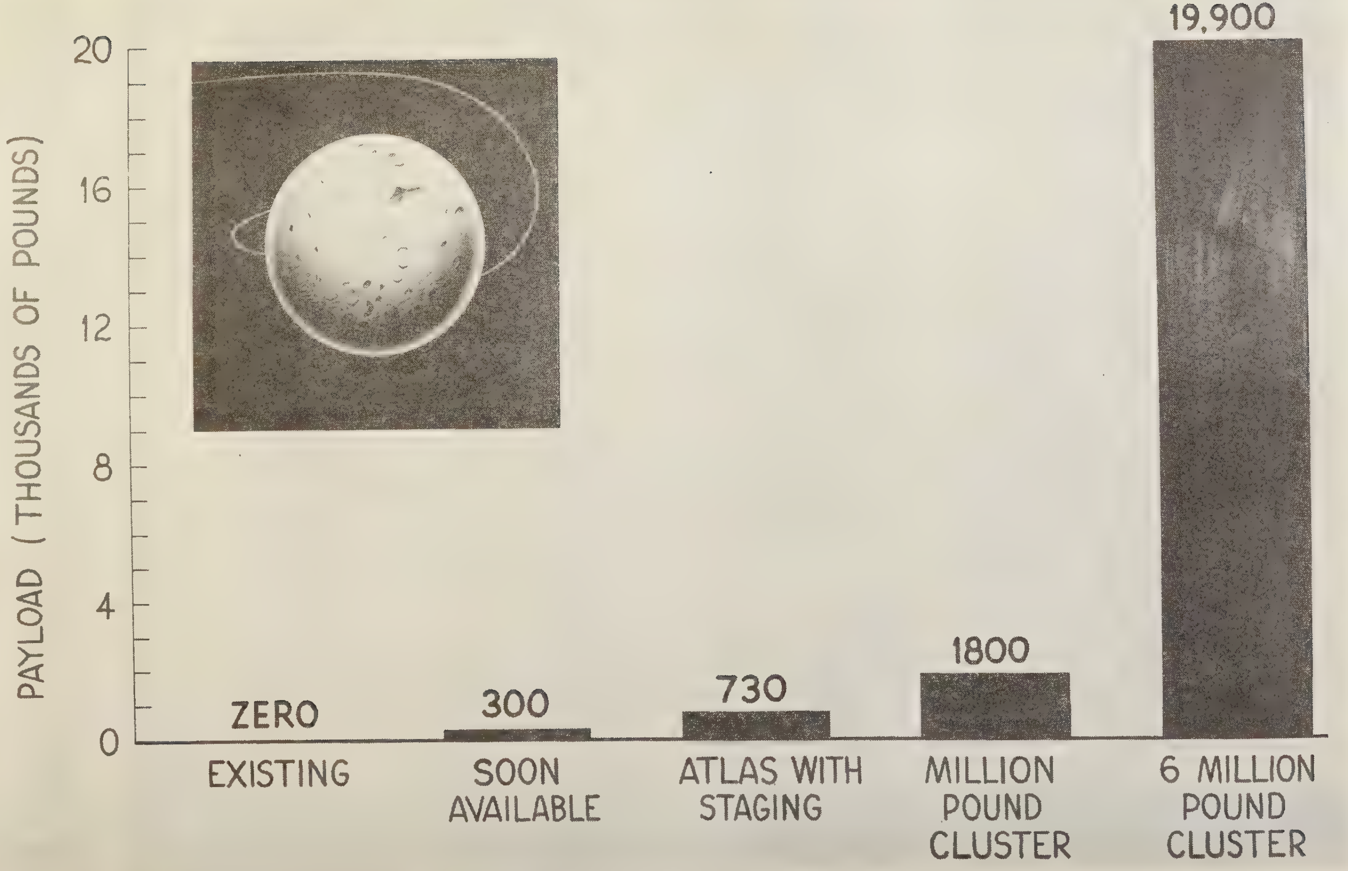


Fig. 7—Payload growth for the lunar landing mission.

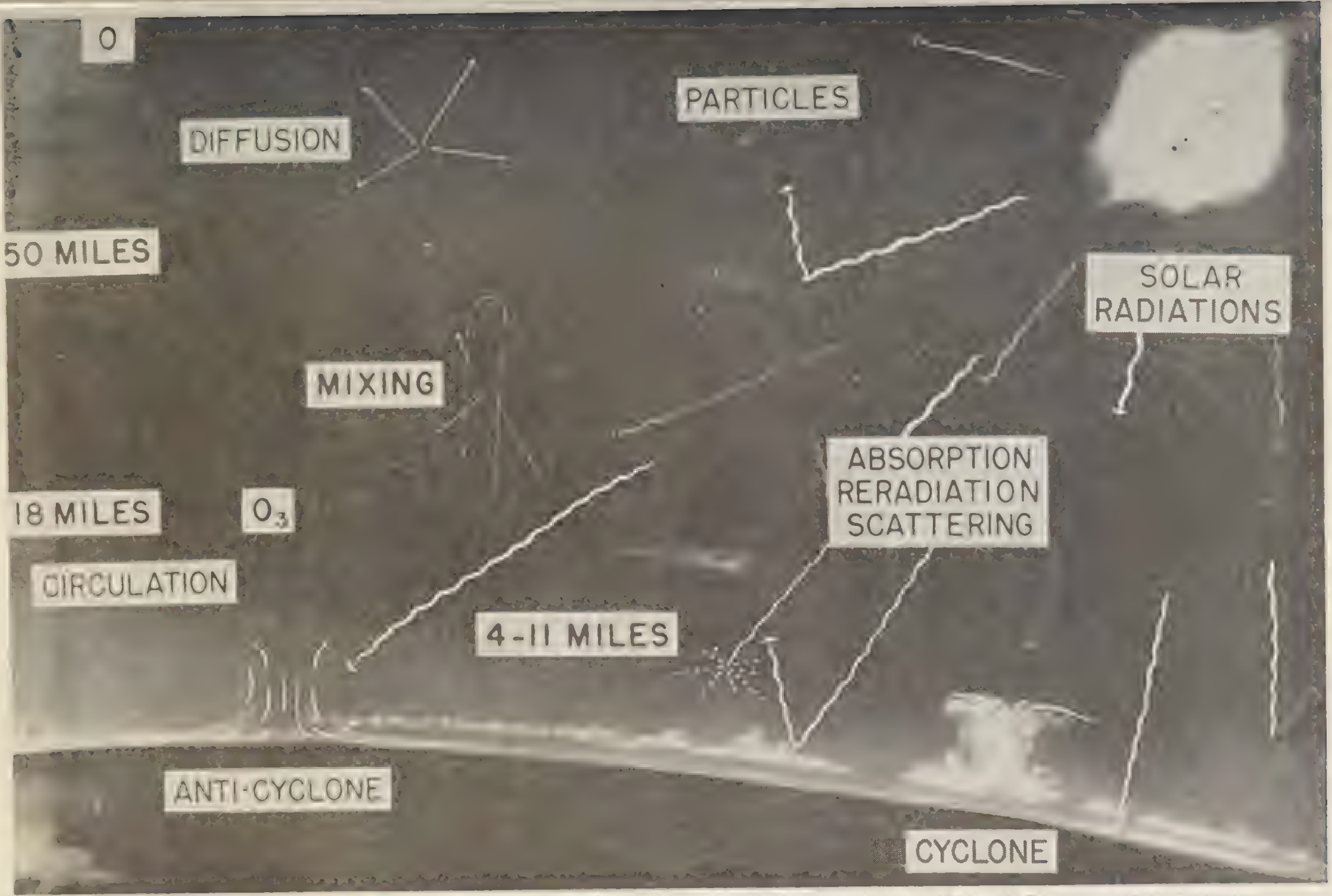


Fig. 8—Atmosphere. Schematic representation of atmospheric phenomena under investigation of the NASA space sciences program

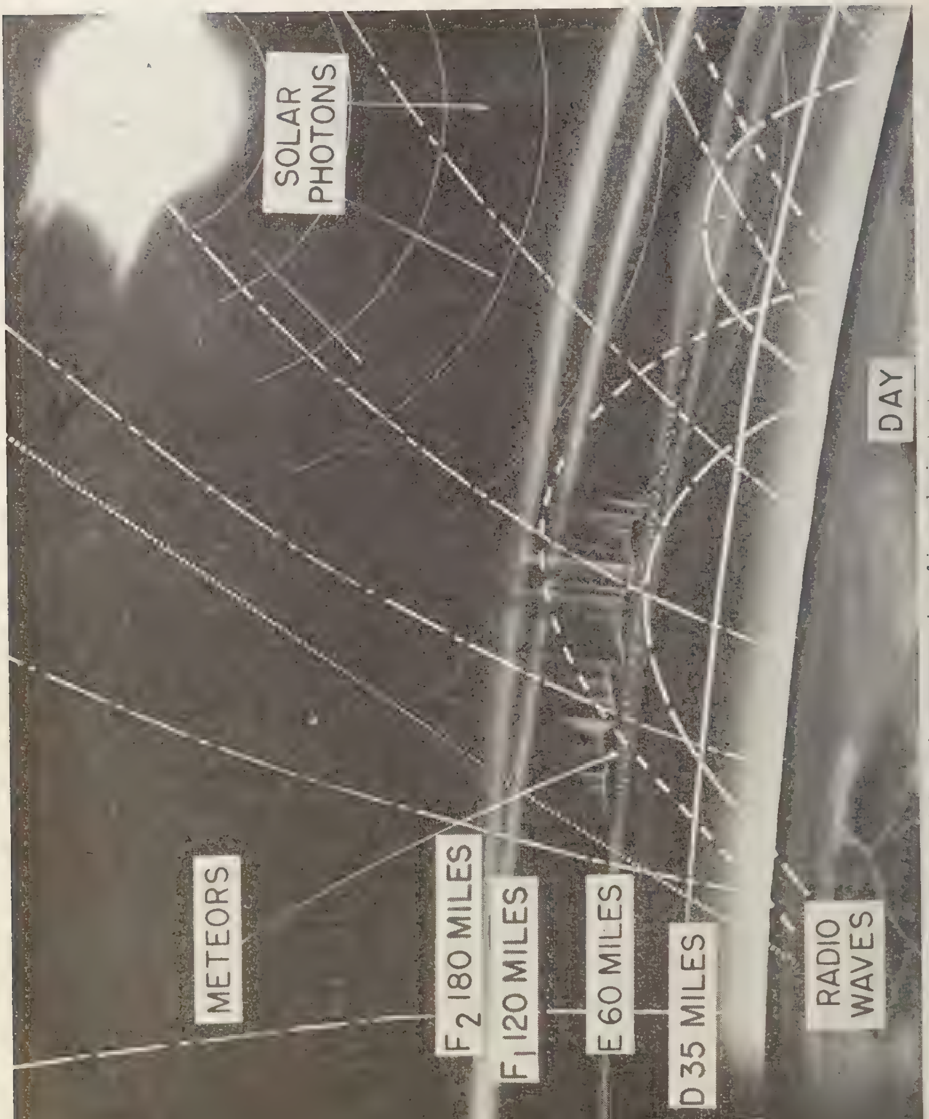


FIG. 9—*Ionospheres*: Schematic representation of ionospheric structure.

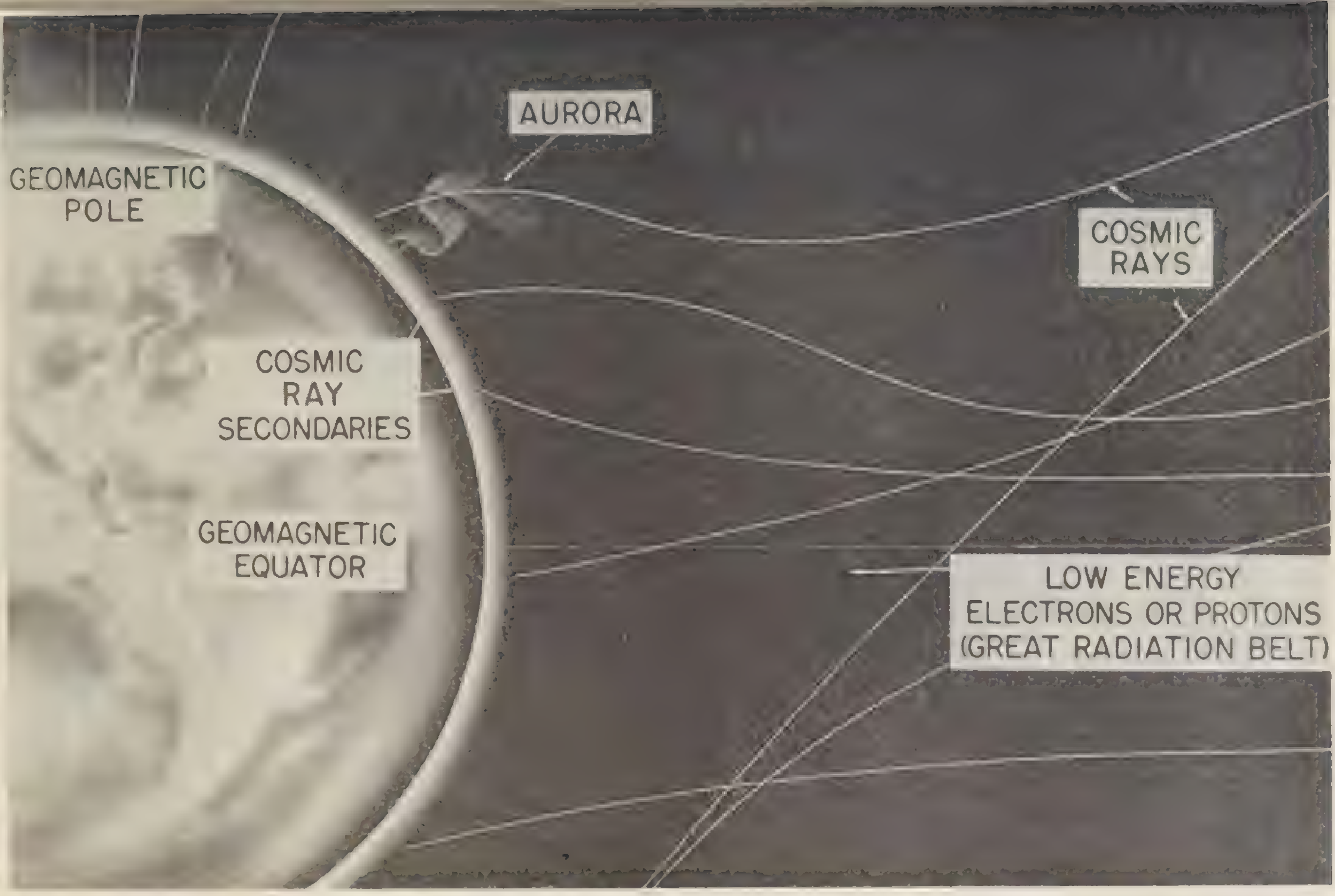


Fig. 10—*Energetic particles:* Schematic representation of the energetic-particle phenomena.

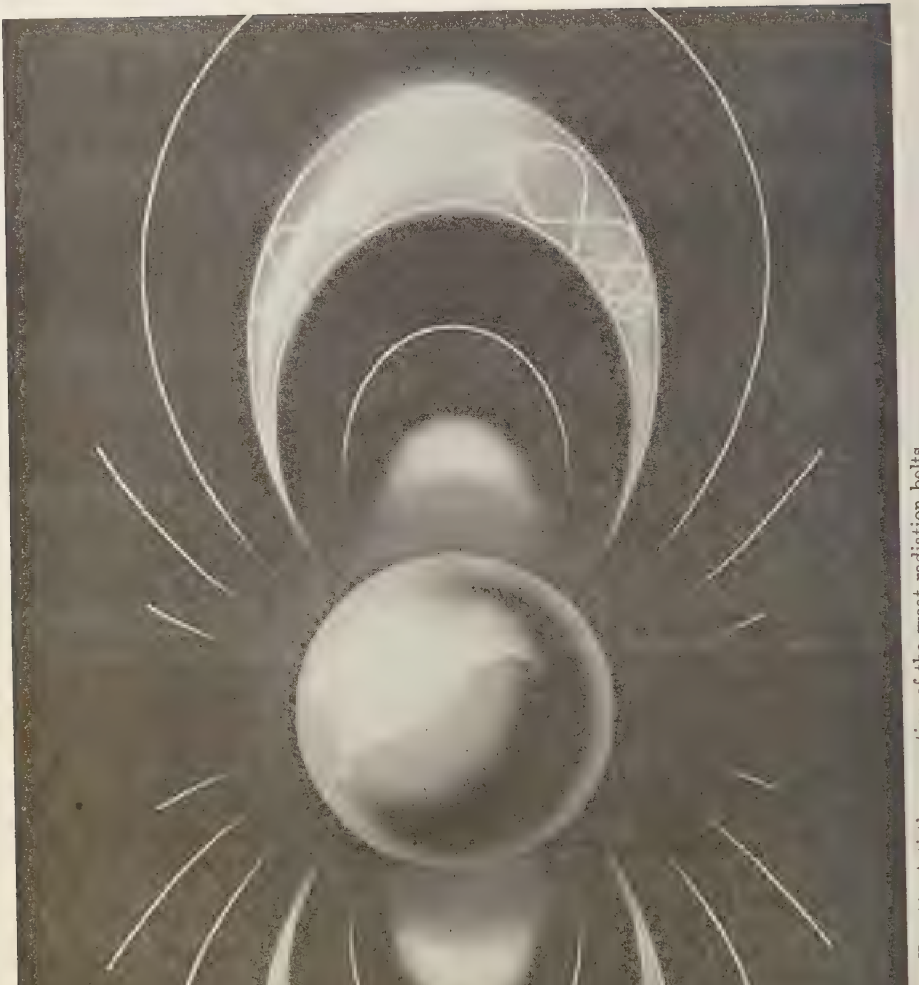


FIG. 11—Artist's conception of the great radiation belts.

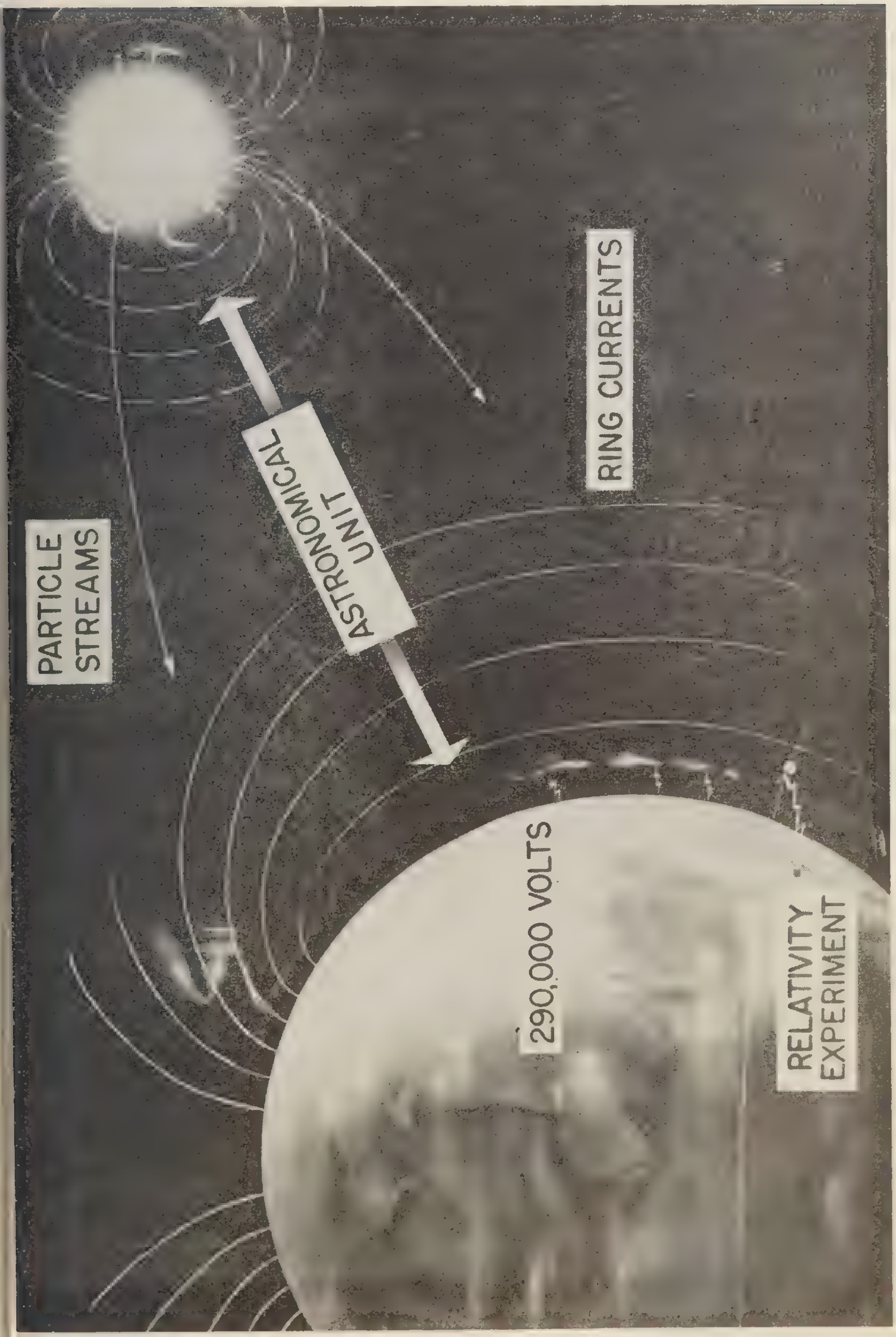


FIG. 12—*Magnetism, electricity, gravity:* Fields of force used for study in the NASA space sciences program.

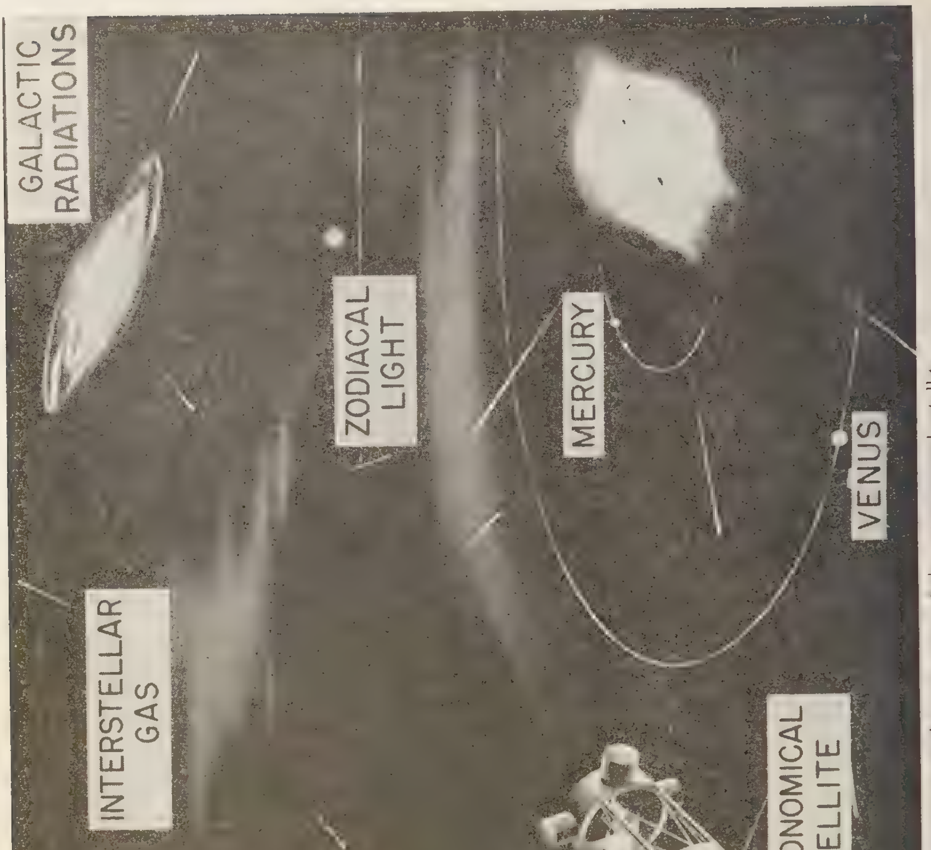


FIG. 13. Astronomy: An artist's conception of the astronomical satellite.

eft. Note the change in scale, also. Just to recall › you the present state of knowledge, during ne IGY we obtained a fairly accurate picture f the ionosphere up to 60 miles, a somewhat ss complete picture up to 180 miles, and only rief glimpses above that level. In fact, the best limpses were provided by the Soviets in their ɔunding rockets and Sputnik III. Again, the 's' ı 'ionospheres' is intentional, because we are ıterested in the ionosphere of the moon, if it as any, and in the ionospheres of Venus, Mars, nd the other planets.

Figure 10 portrays the variety of energetic articles, including cosmic radiation, such phenomena as the Van Allen radiation belt, and ne aurora. It suggests possible relations beveen the aurora and the Van Allen radiation elt. Here again we are interested in whether r not the moon and planets may have such elts.

Figure 11 is a schematic diagram of the great adiation belts. The trajectory shown in the uter belt for a charged particle is actually ıken from one of Störmer's calculations. The adiation belts are a subject of immediate interst, and several payloads are currently under evelopment for flight during the next year r two for study of the energy and density disribution in them.

Figure 12 depicts effects associated with magetism, electricity, and gravity, which are subects of prime interest in the NASA program. ı the near future we hope to put a highly ccurate clock in a 4000-mile orbit to provide ır first controlled check on the general theory of elativity. Work on the clocks is now under way.

Figure 13 shows an artist's conception of an stronomical telescope satellite that will raise quipment above the absorbing and distorting fect of the earth's atmosphere. A near-earth tellite possesses some disadvantages for astroɔmical observations, however. The orbital vecity produces bothersome Doppler shifts, and ne nearness of the atmosphere produces a disırbing amount of back-scattered light. If one ants to look further into the future, it would › easier to use an astronomical observatory on ıe moon. The Doppler shifts would be lower, ıd the fortnight-long night would provide an ;tended period of favorable seeing conditions.

In addition to these research areas, there is now a growing opportunity for biological experiments in the space program.

The success of any program in space research will depend on the interest and participation of the scientific community. Only with sound and ingenious ideas can there be an effective program of observation, measurement, and theoretical research. Part of the NASA space science program will be conducted in the research centers within the agency. In particular, the Goddard Space Flight Center and the Jet Propulsion Laboratory will play considerable roles. But if this is to be a strong national program, the larger part of it must be carried out by other research groups in the country: universities, private research organizations, and industry.

NASA would be pleased to have your ideas and suggestions for a sound space research program and the participation of those who are interested in carrying it out. Suggestions for such research may be directed to the National Aeronautics and Space Administration, attention: The Assistant Director for Space Sciences.

To reduce the labor in preparing such proposals, the initial suggestion should be presented in general terms. NASA members and the proposer may then meet to review the suggested experiment, project, or program to develop a mutually satisfactory arrangement. After such a discussion, it should be possible to prepare a formal proposal that will be satisfactory to both NASA and the proposer and therefore may be accepted without substantial modification.

To ensure that the various space science projects are carried out effectively, NASA is utilizing project working groups, consisting of scientists and engineers actually engaged in preparing for a specific satellite or space probe flight or series of flights. They will not be general advisory groups but will be specifically associated with the project for which they were created and will have the lifetime of that project.

As an example, a working group has been created of scientists and engineers who are participating in the preparation for astronomical satellite observatories. Members of this group will develop instrumentation for galactic observations at various wavelengths. The scientists will be kept informed of the engineering problems associated with fitting experimental equipment into payloads that must match the launching vehicle, endure the vibrations, accelerations, and heating that arise during the launching period, and operate properly in the space environment; the engineers responsible for the solution of these payload package problems will similarly be kept in close contact with the requirements of the experiments to be conducted.

On the international side, NASA will cooperate with other nations in the conduct of space research. NASA has recently offered to launch a satellite for the Committee on Space Research of the International Council of Scientific Unions.

An opportunity has also been provided for those scientists who would like to work for a brief period with the NASA research centers. Both regular and senior research associateships have been established, with stipends starting at $8000 per year, under whose terms men at the postdoctoral level may work either on theoretical problems associated with space research or with experimental groups in the Goddard Space Research Center. The senior appointments have stipends adjusted in accordanc with experience and position of the individua The period of tenure will normally be 1 yea but arrangements for the senior appointment may be made for shorter periods. The selectio from candidates for these appointments will b made by the National Academy of Sciences National Research Council. Those intereste may make application to the Academy.

In closing, may I re-emphasize that the opportunity exists, for those who are interestec to take part in our national program of scientific research and exploration of space. NAS hopes that you, the scientific community, wi take advantage of this opportunity, for it i with your participation, and only with you participation, that the program can be a success. It is only with your participation that thi country can, as the Space Act calls for, maintain a position of leadership in the importan area of space research.

The Moon

Gerard P. Kuiper

Yerkes and McDonald Observatories
Williams Bay, Wisconsin

Empirical data—A brief listing will be given st of the types of information available on e moon. For illustrations of the surface feares discussed in the text, the reader is referred a chapter containing a number of half-tone nar photographs [*Kuiper*, 1959].

May I first comment on the resolving power at can be obtained in studying the surface of e moon. On the best of the existing photoaphs it is 0.4″ (seconds of arc), or 0.4 mile the surface of the moon at the center of the sk.

The visual resolving power in large telescopes, der the best atmospheric conditions and with e mirrors in perfect adjustment, is considerbly better, by a factor of about 4; that is, out 0.1″ or 0.1 mile. This great gain makes it sential to supplement the photographic recds by careful and extended visual observaons. The latter are necessarily time-consuming cause excellent images are obtained only rarely the best mountain climates and with well ade telescopes. Since, however, one good phograph shows more than an observer could acrately put on paper in 1000 hours of observaon, every effort must be made to get photoaphs of the highest resolution; furthermore, otographic records are, of course, objective d precise in location and brightness of surface tail.

The number of signals that can be obtained out the surface of moon by visual observaon is of the order of 10^8, the number of eleents of information in the image being about ^{4} in each coordinate. Each of these elements ay further be observed under different phases illumination. With lower resolving power and her than visual observation different wavegths may be used, and with still lower relving power the two planes of polarization ay be measured in different wavelengths. The total quantity of information that can be so obtained about the moon is very large, much larger than for any of the planets. At an observatory undertaking visual and photographic programs they may well occupy a good fraction of the time during which the moon is placed favorably for observation—which is about one-fourth of the time at any one location (high terrestrial latitudes excepted).

Several of the leading observatories have during the past 30 years made concerted efforts to obtain good photographs of the moon. During the last few years, I have personally examined these collections and secured permission to reproduce the best records and combine them into a photographic lunar atlas. The collections are those of the Mount Wilson Observatory (some 500 plates taken with the 100-inch Cassegrain camera alone); the Lick Observatory (36-inch refractor); McDonald Observatory (82-inch reflector); the Yerkes Observatory (40-inch refractor); the Pic du Midi Observatory in France (24-inch refractor); and smaller collections elsewhere. The *Lunar Atlas* is scheduled to go to press in September 1959. It will be loose leaf and will consist of 200 sheets, each 16 by 20 inches covering 44 lunar areas, each photographed under four different illuminations, together with introductory material. On the prints the diameter of the moon will be 100 inches. Later a supplement will be issued containing special regions on a larger scale and giving also a collection of 'rectified' lunar photographs, i.e., photographs in which the foreshortening toward the limb has been removed by projection on a white globe.

In addition to the visual and photographic records there are several other sources of information that must be considered in developing an interpretive picture of the lunar surface. Among these are the albedo or reflectivity, as well as the polarization of different lunar areas, each measured as a function of the wavelength. The polarization may further be measured as a function of the angle of illumination or phase angle. The polarization curves so obtained may be compared with those of laboratory samples. In this manner it is found that the curves are characteristic of the surface materials; the lunar curves exclude the presence of many substances, reducing the possibilities to powdery silicates like finely pulverized rock or volcanic ash. It is expected that the extension of existing studies to different wavelengths will accurately define the particle size on the moon's outer skin. Current polarization measurements at the McDonald Observatory in the ultraviolet, visual, and infrared have shown the power of this method. Many hundreds of lunar areas can be studied in this manner independently. Such a comprehensive study of the lunar surface is regarded as very desirable before landings are attempted.

Color studies of the lunar surface have also been made. Most areas have been found to have almost identical colors, although the reflectivities may vary greatly. There are exceptional regions, however, which are distinctly yellowish. Examples are the Aristarchus Uplift and an island in Mare Nubium at 26°E, 12°S. These regions appear to be fragments of the old crust that were not flooded by the lavas. More data on the reflection spectrum of these regions are needed before we can hope to interpret the yellow color.

Additional data come from the thermal conductivity of the outer surface layers, which may be determined from thermocouple measurements during a lunar eclipse, and, for the somewhat deeper layers, from microwave data taken during an entire lunation. Radar echoes obtained at different wavelengths gave information on the smoothness of the lunar surface measured in different characteristic dimensions.

Finally, we have information about the moon as a whole, such as its mean density, which is 3.33; the 'figure' of the moon, i.e., the deviation of the lunar body from a sphere (more precisely, the differences between the moments of inertia for the three axes); and the tidal history of the earth-moon system. It is a most interesting fact that the moon is gradually receding from the earth and was, therefore, closer to it in the past. The recession is caused by tidal frictio in the terrestrial oceans. As a result, the eart is losing angular momentum of rotation and th moon is gaining the equivalent amount of angu lar momentum in its orbital motion around th earth, and is thus spiraling outward. There ar indications that the moon was formed quite clos to the earth, perhaps between 0.05 and 0.01 the present distance.

Development of the moon—A study of thes various sources of information, taken in co junction with what is known about the plane and the asteroids, has led to a working hypot esis on the general development of the moo For brevity, this hypothesis will first be state and then used to interpret some of the featur shown by the lunar surface.

The hypothesis is as follows: The solid bod of the moon now observed was formed by process of accretion in a protoplanet or nebu lous mass moving around the sun, before th sun had completed its contraction to its prese small size and resulting great brilliance. Th temperature in this cloud was initially muc below the freezing point of water; there some evidence that the protoplanets separate off from the rest of the solar nebula at a ten perature of some 40°K [*Kuiper*, 1956]. Th earth and moon formed in this protoplanet a binary planet, with a common envelope whi later was dissipated when the sun attained fu brightness [*Kuiper*, 1953]. The collection pro ess or accretion presumably took place son 5.5 billion years ago. Radioactive decay at th epoch was some 10 times more intense than present, and the accreted mass heated up as result, causing partial melting about 4.5 billi years ago. Minor contributory causes to th heating process may have been release of gra tational energy and chemical processes. T timing of 4.5 billion years is based on the dati of meteorites and the earth itself in a series papers well known to geophysicists. Since t asteroids are smaller and the earth is larger th the moon, the approximate agreement betwe the ages derived for these bodies is interpret to mean that this age applies also to the mo itself and, further, that gravitational ener releases have not been dominant (since the ear is much more massive than the asteroids). T conclusion that the asteroids were partly molt d solidified about 4.5 billion years ago is based . much geochemical and astronomical evidence.

The heating led to three types of changes on e lunar surface, all of which may be verified om observation. (1) Large quantities of steam d other gases must have escaped from the terior and altered the nature of the surface aterials from the original loose accreted rub-e to a brittle metamorphic rock. This chance surface texture is apparently responsible for e very different appearances of pre-mare and st-mare craters, referred to below. (2) As the ses were escaping, subsurface explosions may ve occurred which, by analogy with the Kim-rlite diamond pipes in South Africa, may ve led to funnel-shaped craters. As is men-oned below, such craters are actually observed considerable numbers, apparently with the propriate relative age. (3) Actual melting ust have occurred in the interior, which under propriate conditions may have led to the ap-arance of lavas on the lunar surface. We all see that surface lavas have indeed appeared many forms.

After the heating and melting, freezing and oling set in. The cooling and shrinking of sur-ce lavas, resting on a still-hot interior, led to e formation of tension cracks (rills). As would expected, these rills occur only in the maria d in large flooded craters, not in the high-ds. The reduced pressures below the cracks ay have led to renewed melting and the rising new lavas which may have partly filled the acks. Later, as the cooling and shrinking pro-eded to the deeper layers, the filled cracks uld not close and the surface shell could not low the shrunken base below; lateral com-ession would result. Pressure ridges are in-ed present in all the maria and appear to be ated in position to the geometry of the lava sins.

The lunar surface—We shall now review some the principal empirical data bearing on the ocesses of lunar development. Photographs of e full moon show that two main types of sur-e are present: dark areas, or maria; and ghter areas, which are often called the high-ds or the uplands and sometimes the con-ents. 'Continents' seems inappropriate, be-use there is strong evidence that the moon es not have continents in a terrestrial sense, that is, bodies of lower specific gravity, floating isostatically in a mantle.

A closer examination of the bright provinces shows that two subtypes of surface are involved, so that in reality three surface types appear to exist. The two types of bright surface are (1) areas which are bordering on dark areas and appear to be covered with debris thrown out of the dark regions, as by giant impacts, and (2) limited regions, occurring principally in the southwest quadrant of the visible disk, where are found what to all appearances are remnants of the old accreted lunar crust. They seem to fit a common near-spherical surface, which was destroyed locally by impacts but was not obliterated elsewhere. These old areas have probably been the least modified by later events, and, from the absence of near-by lava basins, one would assume that the unmelted crust has remained comparatively thick there. These regions would probably prove to be the softest for purposes of landing, having a surface which might be likened to crusty snow or rather the texture of a rusk; that is, instead of a hard lava surface one would find a surface that would appear stiff only to small pressures and would collapse when larger pressures were applied. The debris-covered regions are best studied under low illumination. They then have a very disorderly appearance, showing that the distribution of the overlying masses is due to a dynamical cause (ejection) rather than a geophysical cause (as lava extrusions from cracks).

There are a dozen maria and a number of smaller mare-like basins. Several of the maria are illustrated in the chapter mentioned before [*Kuiper*, 1959], with accompanying discussion. It may suffice here to draw attention to a few general conclusions.

The maria appear to be of two types: (*a*) Those having considerable symmetry and nearly circular shape (Mare Crisium, Mare Nectaris, Mare Serenitatis, Mare Imbrium, Mare Humorum); these are usually surrounded by mountain rings, and they seem to be nothing but very large impact craters [*Baldwin*, 1949]. (*b*) Maria which appear to be flooded regions; these show no evidence of symmetry and have no surrounding walls, and they have invaded numerous low areas along the shores causing bays, estuaries, etc. Examples are Mare Tranquilitatis (east half), Mare Nubium, and most of Oceanus Procellarum. Among group (*a*), the impact maria, there are still two classes to be distinguished: maria in which the impact appears to have occurred in one phase and the flooding by lava in a later phase (Mare Crisium, Mare Nectarus, Mare Humorum); and those in which the impact occurred at a time when considerable lava masses were available at shallow depths and the lavas splashed over the surface. This appears to have happened in the Mare Imbrium. No such splashes are seen surrounding the first group of maria; we must assume that these impacts were dry, the flooding having occurred later as an upwelling of lava, filling the basin. Mare Serenitatis appears to have formed somewhat before Mare Imbrium, but close to the time of maximum melting, so that its walls are comparatively low and show many regions of subsidence and melting. At the time Mare Imbrium formed, the crust must have regained considerable strength as is shown by the massive Apennines, the Alps, and other mountain groups, including the isolated white mountains between the craters Plato and Archimedes.

It has sometimes been assumed that the lavas of the maria resulted directly from the impacts. No doubt the impacts did cause some heating, but there is convincing evidence that the principal cause of the lavas was the internal heat already stored up in the moon at the time of impact. Pre-mare impacts, even large ones such as caused the crater Clavius, did not lead to any visible melting, and there appears to have been a period approximately coincident with the interval during which the maria formed when even small impacts led to lava-filled basins. Examples are the large number of partly coalescing small basins near Mare Crisium. Post-mare craters such as Copernicus, Langrenus, and Tycho did not lead to any melting either. It appears, then, that several large impacts on the moon occurred during the period of maximum melting, which may be estimated to have been about 1 billion years after the formation of the moon. This result, plus the remarkable concentration of large impacts just before the formation of the maria (as evidenced by the large number of major impact craters outside the mare basins), plus the presence of an old lunar crust which was clearly not the result of accretion by *major* masses, all lead to the conclusio that the large impact craters and the impac maria were formed not as a terminal stage c the accretion process but during a later phas nearly a billion years after the formation of th moon. The writer has supposed that these major impacts were due to the moon's recessio from the earth, which caused it to spiral outward through a ring of small satellites of th earth that had originally formed outside th moon's orbit. Such a satellite ring might b likened dynamically to the formation of the asteroid ring around the sun; such a ring wi form in a nebula surrounding a central mass the nebular density is insufficient for gravitational instability to set in.

In recent years an alternative description c the maria has been publicized, according t which these basins would be filled with a mil of dust. This dust is supposed to have originate from higher areas by a process of radiative an particle erosion. If this view were correct, th floors of the oldest craters would be covere with the same material and would be dar colored, contrary to observation. Only crate roughly contemporaneous with the maria (suc as Plato and Pitatus) have dark-colored floors neither the older nor the younger craters hav them. Furthermore, the maria basins are fa from smooth and featureless; instead, they sho structural detail everywhere of a type incompatible with a thick dust cover. The tension ril and compression ridges in the maria are natur consequences of the evolution of lava basin not of dust; and so are numerous extinct volcanoes, lava domes, sinkholes, and other features observed. These arguments are, howeve but abstract statements compared with the compelling impression generated by direct observation of the moon in a large telescope. It ha been my experience that, where arguments hav failed to convince, the direct view of the moo has succeeded.

The large contrast in appearance of the pre-mare and the post-mare craters strongly supports the hypothesis that the entire moon participated in the heating process. Early pre-mar craters cannot be seen at full moon; evidentl there was no fusion; the surface materials we not altered in texture, but merely displace Post-mare craters, on the other hand, sho ight ray systems or bright halos, demonstrat- g that the lunar surface was brittle during this riod, so that it shattered under the impacts, oducing a whitish rock powder. The crater ys of Copernicus, among others, have, in ad- tion to the white streaks, many scars and uges in the lunar surface, indicating that the ys were produced by spurts of material that ntained a white powdery substance like rock ur as well as larger rock masses that gouged e lunar surface upon impact. Some of the pernican crater rays observed under very low n show the surface there to be quite rough d pitted.

There is much evidence that there are two milies of explosion craters on the moon: im- ct craters, made by infalling bodies which ex- oded upon impact, causing nearly circular aters bounded by walls whose volumes are pproximately sufficient to fill the crater pits; d funnel-shaped pits, of amazing regularity, hose walls are brilliant white at full moon. If e funnels are small, less than perhaps 2 miles diameter, their shapes are inverted cones. t full moon these objects look like brilliant hite circular disks. Craters considerably larger han 2 miles in diameter have slopes similar to e small ones, approximately 30° with the hor- ontal, but are truncated cones, whose bottoms e darker than the sloping walls, matching ap- roximately in albedo the surrounding lunar rface. At full moon these objects look like illiant white rings, resmbling washers or the ing Nebula in Lyra. Studies of these objects ow the walls to deviate slightly from precise nes, being slightly convex. This is best demon- rated by comparing the lunar craters with odels cut out on a lathe, appropriately painted d illuminated. The shapes are quite different om the bowl-shaped impact craters, and it is erefore assumed that they were not formed meteoritic impact but by lunar explosions. he fact that these inverted cone craters may cur in crater rows (which seem to be rows explosion pits, possibly situated above a dyke, was pointed out to me by Professor H. H. ess) strengthens this interpretation of the ater cones. It is remarkable that these ob- cts are often nearly rimless; if rims are pres- t at all, they are low and very smooth. Ap- arently the materials were finely divided and hurled to great distances, so great that not even white halos appear around these craters. Explosion craters of this type occur both in the maria and the highlands.

We have referred to the appearance of lavas at the lunar surface through impacts during the period of maximum melting, and also to associated floodings in other regions. One of the best arguments that actual lavas were present is based on the presence of craters like Fracastorius on the shore of Mare Nectaris, whose walls on the side of the mare collapsed and merged with the mare itself. Additional evidence for lavas is seen from several groups of extinct volcanoes, particularly prominent in the field just east of Copernicus; in a crater like Wargentin, which appears to have been filled with lava from below to the point of overflow; and probably in the central mountains of impact craters that formed just before the maria. The oldest impact craters on the moon have no central mountains; post-mare craters have comparatively small and broken-up central mountains which are brilliant white at full moon and may be composed of piles of debris. Craters formed during the period of maximum melting have, on the whole, flat lava-colored crater bottons; Plato, Archimedes, Ptolemaeus, and Pitatus are good examples. Only the late pre-mare craters appear to have prominent central mountains. This is interpreted to be due to subsequent upwelling of lavas through cracks left in the crater bottoms, the impacts having occurred after the surface layers of the moon had already become brittle. This interpretation of the central mountains is strengthened by the observation that the deepest craters in this class, like Albategnius and Arzachel, have the highest central mountains. In some craters a good part of the floor is covered by mountains, large and small. In others, such as Pythagoras, even a major mountain may be appreciably off center. In still others, cracks have formed in the floor which appear to have given rise to a row of mountains, covering the entire radius. Finally, lava extrusions appear to be very common on the moon. There are many rounded hills in the neighborhood of crater Schiller, which seem to be extrusions, since they often have riding on their crests narrow, dike-like, smaller extrusions. Then there are several mountains south of the

Carpathians and just West of Mare Humorum which have a distinctly linear structure, apparently due to the presence of tension rills from which extrusions developed. Finally, many pressure ridges appear to be cracked open near their crest and in several instances lavas appear to have been extruded from these rills.

We have referred to the production of tension cracks (rills) and pressure ridges as secondary phenomena accompanying the cooling and shrinking of the lavas in the mare basins. The pressure ridges commonly exhibit en echelon structure which, on the basis of geophysical experience, indicates the dynamical interaction of a cool and buckling upper crust with a shrinking deeper layer having a different structural direction. Good examples are the Serpentine Ridge in Mare Serenitatis [*Kuiper,* 1959, Fig. 11), the ridge system south of the Rainbow Bay [*op. cit.,* Fig. 14], and the ridge system south of Marius [*op. cit.,* Fig. 19] The pressure ridges and the lunar volcanoes are presumably the most recent phenoma resulting from the igneous origin of the mare basins; but they are probably not much less than 4 billion years old, in view of the cooling rates of the outer layers of the moon, which can be estimated. Also, the volcanoes east of Copernicus have been covered by crater rays ejected by Kepler; none of the extinct volcanoes give any evidence of being comparatively recent. In this connection I should perhaps state my views about the recent announcement of possible vapors on the floor of crater Alphonsus [*Alter,* 1957]. Inspection of the published photographs indicates that the blue plates were taken with much less resolution than the red plates, not only within crater Alphonsus but also on its rims and outside it. The differences, blue–red, appear to this writer as due to seeing differences, not to obscuration of surface detail. The absence of a lunar atmosphere and the continual bombardment of the lunar surface with energetic solar particles and radiations make it extremely unlikely that escaping gases could accumulate during the lunar day, as the alternative interpretation of the photographs would require. If an empirical check were desired, this could be made from polarization measurements, since the radiation scattered by gas would be nearly completely polarized at this illumination.

The last phase of lunar development is stil active: (1) the bombardment by meteorites which appears to have led to the formation o many thousands of visible meteor craters, mos of them less than a mile in diameter, and al showing bright halos; (2) the bombardment b cometary debris, which gives the lunar surfac some of its photometric properties such as th absence of limb darkening; and (3) the exposur to solar radiations of all wavelengths as well a to streams of solar particles. On these small scale phenomena much remains to be learne both from observations and from laboratory experiments.

In the years to come, the exploration of th moon by rockets promises to add new and vita information about the composition and structure of the lunar surface. The great efforts and expenditures that these experiments will demand make it advisable to intensity the lunar studies from ground-base observatories with supporting laboratory work. I estimate that no more than 10 per cent of the significant information accessible to inexpensive ground-based studies has so far been obtained.

References

Alter, D., *Pubs. Astron. Soc. Pacific, 69,* 158–161 1957.

Baldwin, R., *The Face of the Moon,* University of Chicago Press, 1949.

Kuiper, G. P., *La physique des comètes,* chapter 31, Louvain, 1953.

Kuiper, G. P., On the origin of the satellites and the Trojans, *Vistas in Astronomy,* vol. II, Pergamon Press, New York, 1956.

Kuiper, G. P., The exploration of the moon, *Vistas in Astronautics,* vol. II, Pergamon Press, New York, pp. 273–312, 1959.

Discussion

Question: Is it true that some of the other maria in addition to Mare Imbrium also present radial and concentric structures, i.e., fracture systems and radial valleys which resemble some of the characteristics of Mare Imbrium?

Mr. Kuiper: I think that the Imbrium impact was found to be the largest event on the front of the moon. There is a system of similar structural lines in the upper part of the picture, near the south pole of the moon, which cannot be attributed to Mare Imbrium or to any other feature on the front of the moon.

By studying the convergence of these structural lines, one is inclined to postulate a basin somewhat similar to Imbrium and possibly somewhat rger on the back of the moon, at latitudes 18° S nd 32° behind it.

There is some evidence to support this hypothsis, namely, that some of the very highest mounains on the limb of the moon may well be the eripheral range around this basin.

But otherwise Mare Imbrium appears to be the nly one that has such a prominent system of tructural features associated with it.

I believe that the other maria were made arlier. Imbrium was apparently the last of them.

Question: What is the significance of the white adiating lines from what appeared to be the pole the first slide, near the upper right-hand corner?

Mr. Kuiper: One finds that these ray systems ad a variety of sizes. The Tycho rays are the ongest. I have come to the conclusion that the ays are caused by spurts of material thrown out f the impact crater and containing two types of aterial: they contain a white material, a sort f rock flour, which causes the visible rays; they lso contain large blocks of rock that cut through he lunar surface and cause the gouges.

Question: Wouldn't it be possible with known techniques to make a topographic map of the moon which is based on available data?

Mr. Kuiper: That is a question we have considered rather carefully. The base points on the moon are not sufficiently accurate at the moment for a good topographic map. But precision regional maps based on the length of shadows can be made.

We also have started a program of base point determinations. But at the moment the vertical coordinates of the base points are uncertain by something like a kilometer and therefore not accurate enough for purposes of photography.

Question: You identified round smooth objects as extinct volcanoes. In none of them could I detect any sign of a small crater.

Mr. Kuiper: This is where the visual resolving power of the eye enters. Visually, in all these objects that were pointed out, as volcanoes, the calderas can be seen very clearly, but not in the photographs.

Primary and Secondary Objects

Harold C. Urey

University of California, La Jolla, California

In considering our space program, it is well to study objects that arrive from extraterrestrial sources quite without any effort or expense on our part. It is my purpose to discuss recent and older observations of meteorites and to draw some tentative conclusions from them.

Meteorites are objects of variable structure and chemical composition, and a complete review of their properties is impossible in a brief time; therefore, this discussion will be limited to some specific features from which certain conclusions will be drawn. I believe that these conclusions will not be contrary to evidence that is not reviewed.

Briefly, two sets of objects, the primary and the secondary, are required to account for the properties of chondritic meteorites [*Urey*, 1956]. The primary objects are of about lunar size, and it is suggested that the moon is one such object. The secondary objects have been identified with the asteroids in my past publications, but they may possibly be the surface regions of the primary objects; in fact, the surface of the moon may be the immediate place of origin of the stone meteorites.

Iron and stony-iron metorites—Prior's catalogue [*Prior and Hey*, 1953] lists 12 stony irons and 42 irons that have been observed to fall. Iron meteorites consist of a metallic alloy mostly of iron and nickel, nickel being present in amounts ranging from about 6 to 60 per cent, but usually about 7 or 8 per cent in the iron meteorites. Both the kamacite (body-centered cubic) and the taenite (face-centered cubic) crystal modifications are present, and when polished surfaces are etched these appear as the Widmannstätten figures. As *Henderson* [1958; private communication] has observed, the surfaces of the objects are often deeply pitted, with holes sometimes 8 times as deep as their diameters, and they could hardly have been formed by the breakup of larger objects. The stony irons consisting of mixtures of silicate and metal in roughly equal proportions must represent some boundary regions between metal and silicate phases. In fact, some specimens of the Brenham Township pallasite consist of typical iron of the octahedrite classification attached to a typical specimen of a stony iron; hence the boundary region is preserved in them. These facts argue that metal masses of the approximate size of the metal meteorites were embedded in silicate objects: a 'raisin-bread' structure for the parent object rather than a planet with a large core. These arguments of Henderson seem very reasonable. The objects representing surface regions between the silicate and metal phases are numerous, and hence the surface regions must have been large and the metal regions small. Again, there is evidence that differentiation into a planetary core in the primary object did not occur.

Because of their importance in understanding the chondritic meteorites, some of the metal particles in the Brenham Township pallasite are shown in Figure 1. This specimen was not well polished, and the diagonal streaks should be ignored. A considerable variety of kamacite (K) and taenite (T) particles can be seen in this section. The taenite particles range from those having a smooth gray appearance to those with a dark interior and marked diffusion borders. These particles were formed by diffusion of iron and nickel between the two crystal modifications in accordance with the requirements of the iron-nickel phase diagram (Fig. 2).

The differentiations took place by cooling from a higher temperature along the line *AB*. At *C*, kamacite of composition *D* appeared, and as the temperature fell, more kamacite was present, and less taenite, until at 450°C the compositions would be those of the points *E* and *F*, and the ratio of kamacite to taenite would be *BF*/*BE* if equilibrium was attained. The rate

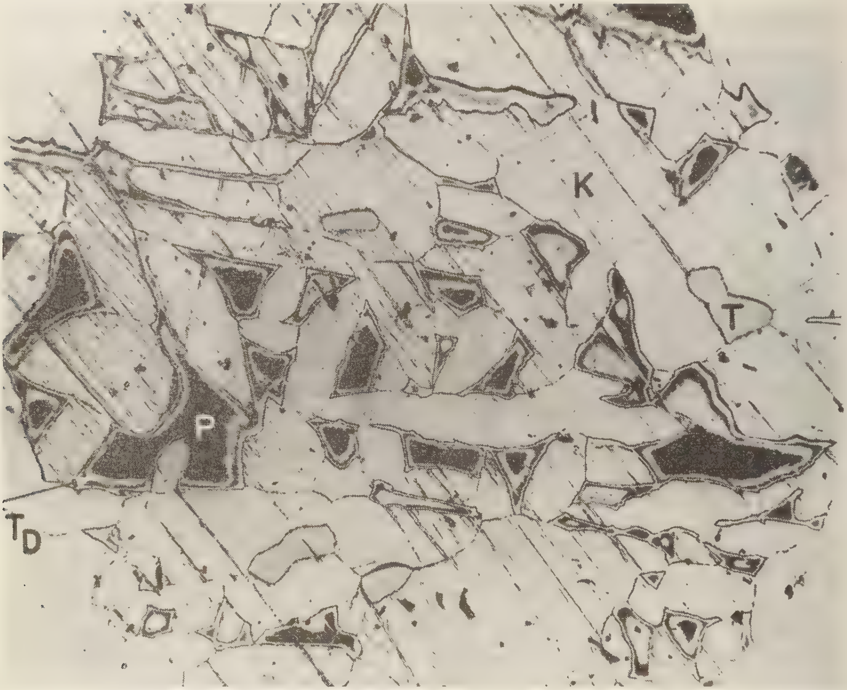


Fig. 1—Brenham Township. Note that the taenite particles are sometimes clear and sometimes have a diffusion border with a plessite interior. Such particles are found in the stone meteorites.

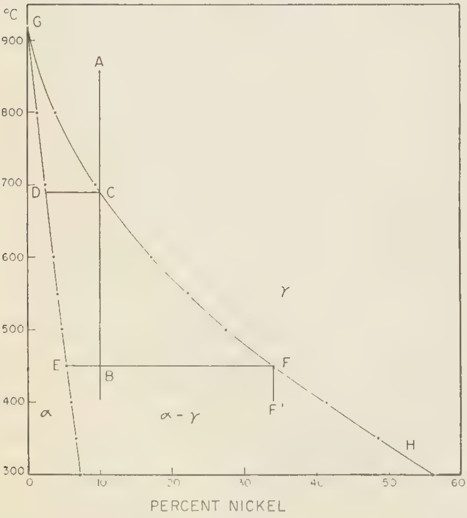


Fig. 2—The Fe-Ni phase diagram.

of diffusion would become much slower at lower temperatures until diffusion from the interior of the taenite crystal would cease, and thus a taenite particle of composition *F* would move to the point *F′*, where it would be unstable and kamacite would form within the taenite particle. The kamacite etches to a dark color, giving the dark interior of some taenite grains, shown in the illustrations. The mixture is called plessite. These relationships are well known [*Perry*, 1944]. They are emphasized here only to show that we do understand how the metal particles of some stone meteorites must have been formed within masses of iron-nickel similar to those of Brenham Township. The temperatures are estimated to have been some 400 to 450°C, and the times required some tens of millions of years. We envision a mass of iron-nickel embedded in a silicate matrix cooling slowly for long periods.

Stone meteorites—Stone meteorites are classified into many groups on the basis of structure

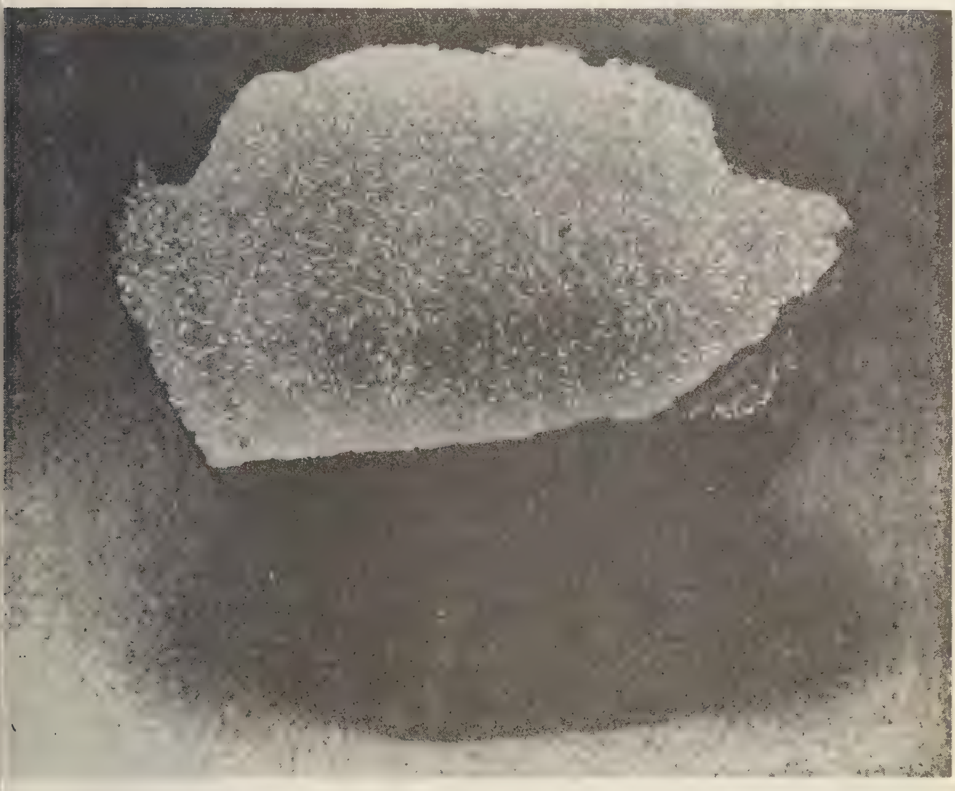


Fig. 3—Beddgelert meteorite. This plate shows the conglomerate structure of a typical chondritic meteorite.

and mineral composition. Chondrites contain from few to many rounded silicate objects called chondrules; achondrites contain very few or none at all. Stone meteorites are conglomerates, in general, though the chondrites show this feature more prominently. The structure is shown beautifully by many pictures of polished surfaces of meteorites; the Beddgelert meteorite photographed by Paneth (Fig. 3) shows it very well. Obviously the meteorite did not solidify from a melt, for the metal particles even in a weak field would not have remained uniformly suspended throughout the mass. All chondrites are very similar in this respect. In addition, all are porous, some markedly so.

The metal particles of all chondrites examined so far consist of both kamacite and taenite, and very often the two kinds of particles are separated from each other and are completely surrounded by silicates. In some chondrites the taenite particles are clear, and in others the interiors contain plessite and have diffusion borders just as is observed in Brenham Township. All the metal particles are distorted more or less, and in a few specimens the taenite particles are broken off[1] (see Figs. 4 to 8).

A study of the plates shows that some violent crushing process broke the original mass into small particles, which were accumulated into a conglomerate mass and compacted and welded into a strong object. This object became the immediate parent of the chondritic meteorites. Sometimes fragments of one variety of stone meteroites are found within another of quite a

[1] The metal particles will be discussed in detail in a paper to appear soon in *Geochim. et Cosmochim. Acta* by H. C. Urey and T. Mayeda.

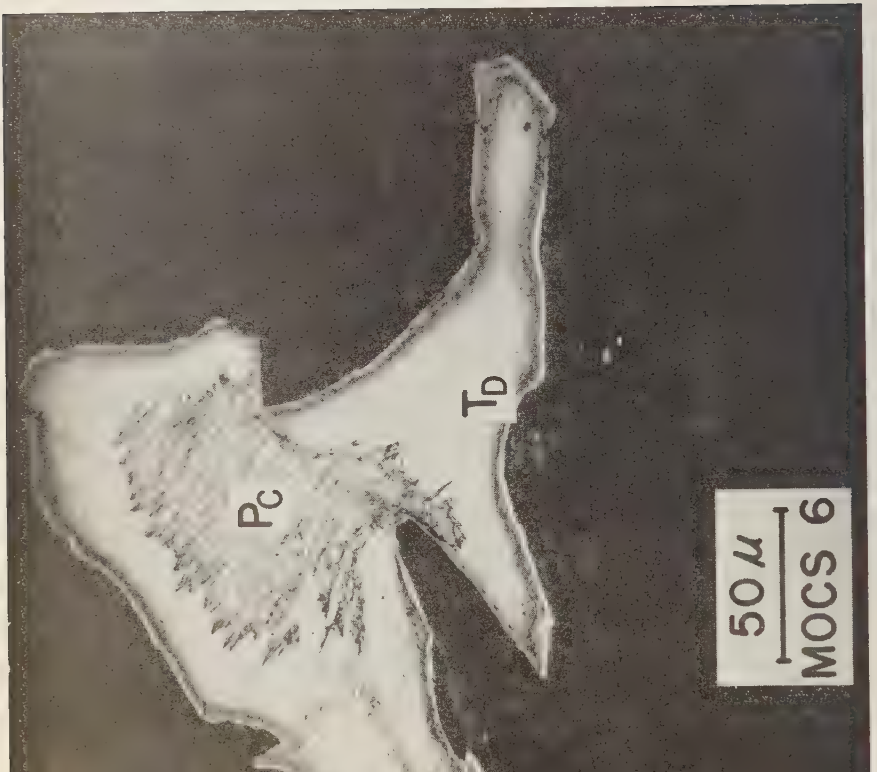
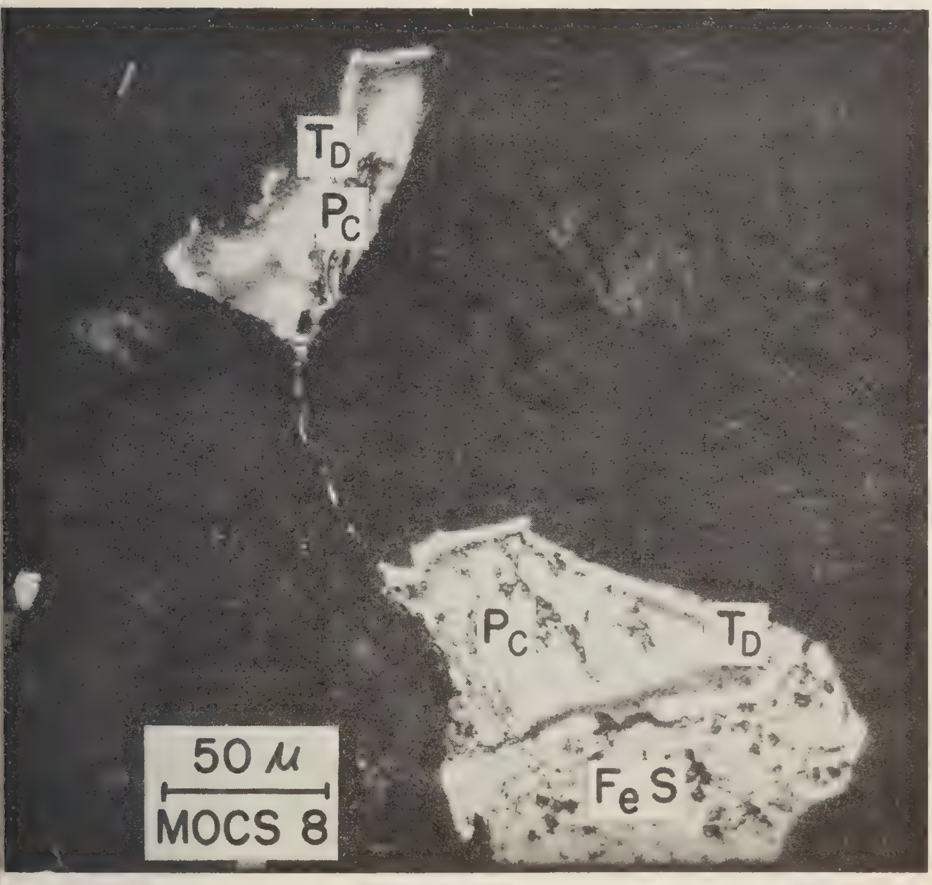


FIG. 4—Mocs. This shows a taenite particle which has been broken. Note the diffusion border and the plessite interior, and note also that the particle is completely surrounded by silicates which photograph black.



ı. 5—Mocs. Another broken taenite particle. Note the connecting line of metal. The two pieces have not been separated, and hence crushing without scattering has occurred.

ferent classification, e.g., a black chondrite
the white achondrite of Cumberland Falls.
ıat is required apparently is a great sandbox
ere the primary object is broken up and com-
cted. The compacted object, in turn, is broken
l scattered into neighboring boxes where some-
at different materials are being processed.
e rubble is again compacted and redistributed.
this way the origin of the so-called polymict
uctures [*Wahl*, 1950] can be understood.

The chondrules are sometimes glassy even
w and contain particles of metal of very small
aensions. They apparently were formed by
ezing liquid drops [*Sorby*, 1864]. Other chon-
tes are crystalline and look as though they might have been formed by crystallization of liquid drops, though the evidence is not clear. A great puzzle is posed by the presence of both liquid drops and taenite particles with their diffusion borders. If the chondrules were made before the metal particles, why did they not crystallize during the cooling process that produced the metal particles, and why were they not broken during the breakup of the metal masses? If the metal particles were made first, why were the detailed patterns in the taenite particles not destroyed during the heating process that produced the melted silicate? Somehow the drops of silicate and the metal particles were produced separately and then mixed intimately together.

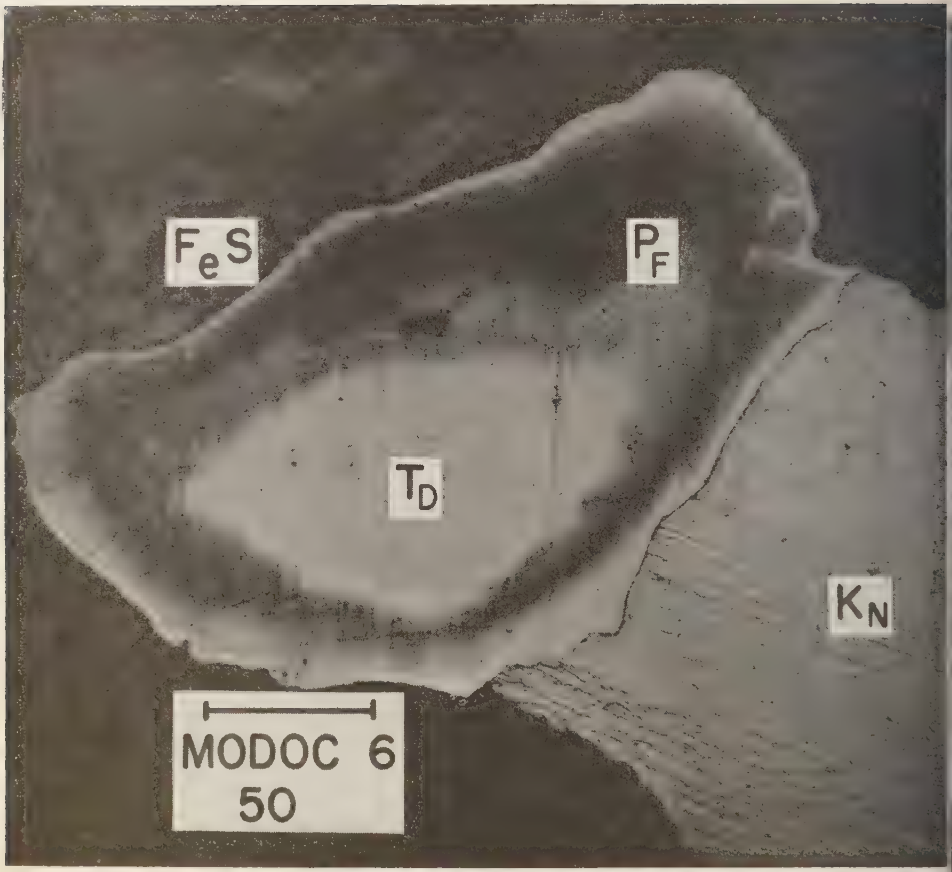
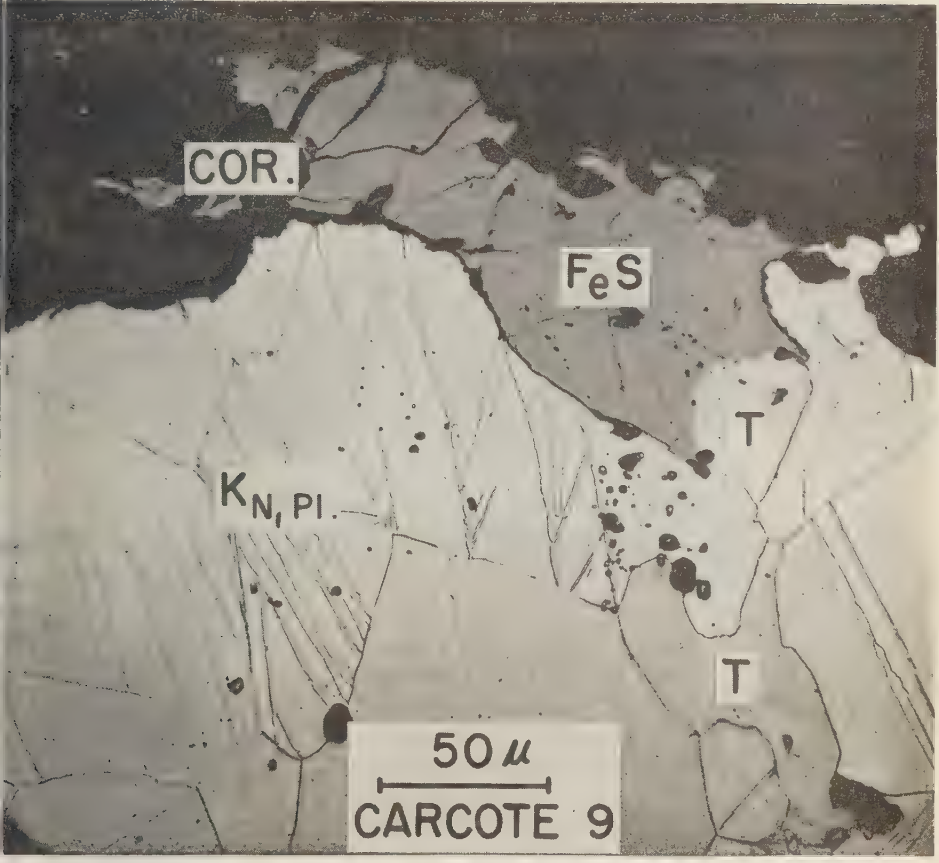


FIG. 6—Modoc. A taenite particle is surrounded by FeS, kamacite, and silicate, and yet all the diffusic borders are very similar. It could not have been formed in its present location.

Was it in a vast cloud above the surface of a primary object during some collision phenomenon? Collisions would appear to be the only process for breaking up the metal aggregates. We note that great collision processes have occurred during the formation of the lunar maria.

Diamonds are found in a number of iron meteorites, and pseudomorphs of diamonds have been reported also. Two stone meteorites, Novo Urei [*Jerofejeff and Latschinoff*, 1888] and Goalpara [*Urey, Mele, and Mayeda*, 1957], contain diamonds. It seems likely that diamonds can be made only at high pressures. No other high-pressure minerals have been reported, but nature is most uncooperative. The irons could not contain high-density silicate minerals; Goalpara contains only 0.04 per cent of sodium, an
no aluminum is reported, so that jadeite cann
be present. The analysis of Novo Urei is r
ported to show 0.60 per cent Al_2O_3 and no alk
lies. Other stones have been examined, but r
diamonds have been found. We assume that th
diamonds could be formed only under hig
pressures such as are present deep in the moo

The ages of meteorites are determined b
three methods. The lead-lead age [*Patterso
Tilton, and Inghram*, 1955] measures the tim
since the iron meteorites were separated fro
uranium and thorium, presumably by a meltir
process followed by separation of silicates an
metal in a gravitational field. The rubidium
strontium method measures the time since rub



'IG. 7—Carcote. Polycrystalline kamacite with two taenite particles included in it. In this case dif- entiation *in situ* might have occurred. Yet the edges of the metal particle are consistent with a breakup of a larger particle.

.m was removed from the Pasamonte achon- te and its strontium was increased in amount :*humacher,* 1956]. The potassium-argon ages '*asserburg and Hayden,* 1955] measure the ıe since stone meteorites last lost their argon. ıce some diffusive loss of argon is probable, ɪse last ages are probably too low. All these asurements indicate that the events we are ısidering took place some 4.5 × 10^9 years ago.

Model for the origin of the meteorites—About aeons ago, then, objects of lunar size ac- nulated in which a heating process occurred der such conditions that slow cooling during ne tens of millions of years followed. (It is t my purpose to discuss this process. It is not easily devised, since objects of lunar size cool very slowly.) These primary objects were probably broken up by collisions and reaccumulated into some secondary objects which, when further broken up, supplied the chondritic meteorites. No simpler process seems feasible for accounting for the facts.

The moon may be one of these primary ob-objects, as I realized after devising what seemed to me a reasonable model for the *grandparents* of the meteorites. Some eight objects having masses within a factor of 2 of the lunar mass exist in the solar system now, and the next satellite in mass to these has only 3 per cent of the lunar mass. Since in order to break up an

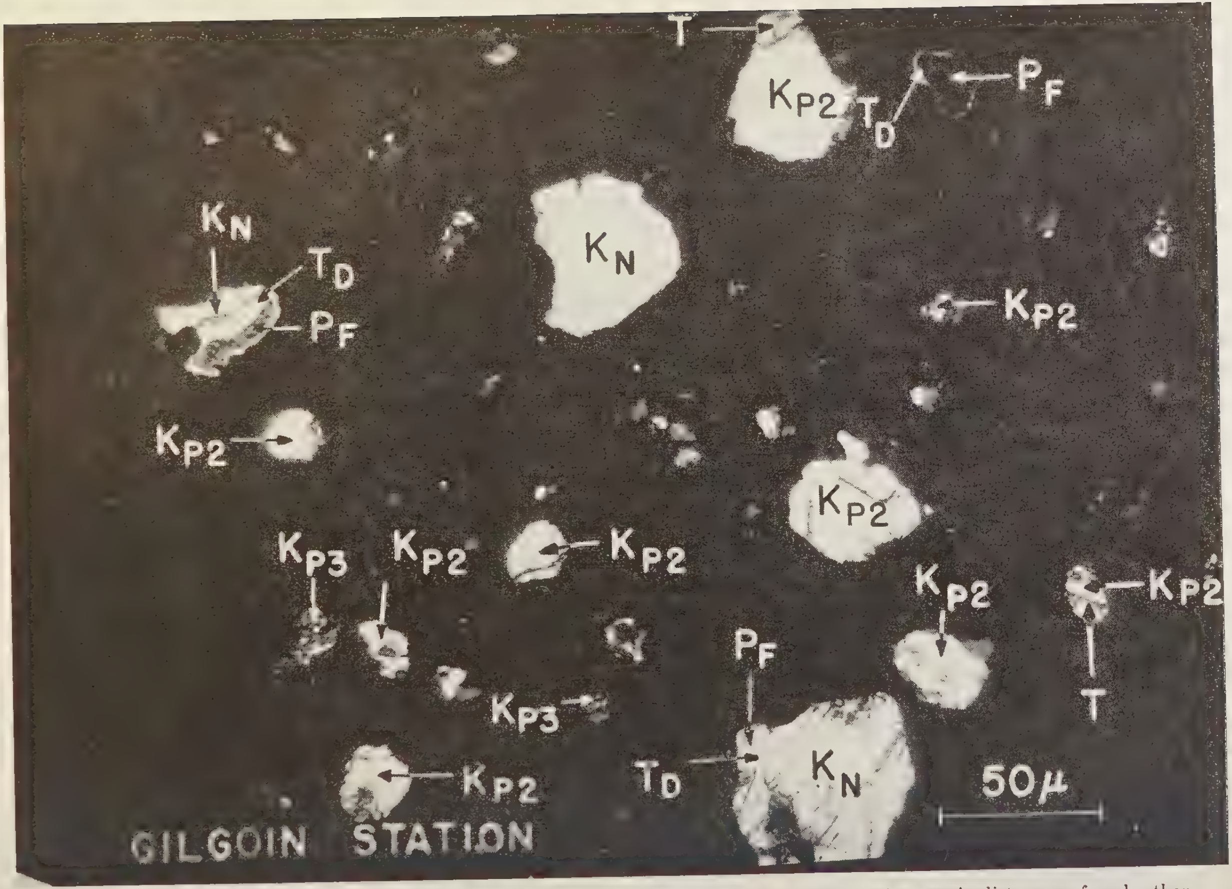


Fig. 8—Gilgoin Station. Note the many metal particles of complex composition within microscopic distances of each other.

TABLE 1—*Comparison of the density of the moon with calculated densities of meteoritic matter*

	Observed	Calculated
Part 1		
. Density of low-iron-group chondrites	3.51	3.574
. Density of high-iron-group chondrites	3.66	3.761
. Low-iron group with albite converted to jadeite and SiO_2 in $MgSiO_3$		3.653
′. Density of moon at low temperature and pressure: $\beta = 7.9 \times 10^{-7}$, $\alpha = 3.3 \times 10^{-5}$, $t = 1100°C$ (olivine)		3.41
. Required iron content of (3) in order to have density (4)		
a. Iron present as Fe and FeS		10.78
b. Iron present as FeO		11.52
. Cosmic abundance of Fe (Si = 10^6) (5*a*)	2.44×10^5	
(5*b*)	2.65×10^5	
Part 2 (Using other constants)		
′. Density of moon at low temperature and pressure: $\beta = 10.1 \times 10^{-7}$, $\alpha = 2.76 \times 10^{-5}$, $t = 1100°C$ (enstatite)		3.382
′. Required iron content of (3) in order to have density (4′), removing FeO and FeS		9.11
′. Cosmic abundance of Fe (Si = 10^6)	1.95×10^5	
. Suess-Urey abundance of iron	6×10^5	
. Aller solar abundance	1.4×10^5	

* The values of α and β are those for forsterite and are taken from *Geol. Soc. Am. Spec. Papers, 36, 33* d *36*.

† The value of β is that for enstatite, and the value of α is the mean of augite and diopsite taken from *ol. Soc. Am. Spec. Papers, 36, 32* and *57*.

oject of lunar mass by collision it is necessary at the colliding object have a similar mass, a tal of ten objects is accounted for in this way. ich objects may have been far more numerous the past than they are now and may have en important during the evolution of the solar stem. In any event, these objects may well ve been composed of the more nearly correct lar average material of the less volatile kind an the earth and other terrestrial planets.

Table 1 gives the results of some estimates the proportion of iron in the moon. The nsity of the moon is 3.34 at known pressure d unknown temperature. An estimate, based several models, for the mean lunar temperare is 1100°C with possibly an error of ±200°C. e need to know the coefficients of expansion d compressibility in order to calculate the ean density at low pressure and ordinary temratures. The table shows two sets of values lected and the densities so calculated: 3.41 and 3.382 g cm^{-3}. These densities are lower than those of the chondritic meteorites, of which there are two well defined groups [*Urey and Craig*, 1953]. The observed densities are less than those calculated from their normal minerals, probably because of experimental errors, since voids in these porous objects are common.[2]

In the calculations it is assumed that all albite in the moon has been converted to jadeite and $MgSiO_3$. If so, the density of the moon must be comparable to the density of the low-iron-group chondrites given under item 3 of the table. The iron content of the moon appears to be very low and to agree approximately with more recent estimates for the sun [*Aller*, 1958]. If high-density minerals are not assumed to be present

[2] These results differ from those calculated by Urey and Craig because they used an erroneous density for $FeSiO_3$ taken from handbooks of physical constants.

in the moon, the abundance of iron is about 14 per cent. If some 2 per cent of water is assumed to be present, the composition of the moon is the same as that of the low-iron-group chondrites. However, this amount of water is much greater than that observed in the meteorites and the earth. If such a large amount of water is present in the moon, the surface rocks of the moon can be expected to contain appreciable amounts of water. Lunar explorations may answer this question in the near future.

We turn to the question of the identity of the secondary objects. It has been mentioned above that the surface of a primary object would solve some problems of the origin of the chondrites. It is usually thought that the meteorites came from the asteroids inasmuch as some objects of this class do cross the orbit of the earth, and approximate orbits of meteorites have been calculated that agree with the orbits of the asteroids. It is a view to which I have subscribed, and I do not wish to abandon this possibility definitely.

There is evidence that the stones and irons generally do not come from the same region of space, and that they definitely have quite different histories. Table 2*a* and *b* gives the cosmic-ray ages of the stone meteorites as determined from their He^3, Ne^{21}, and A^{38} by the observers listed at the end of the table. In Table 2*a* are given the results for meteorites for which both H^3 and He^3 have been determined, so that the shielding from cosmic rays because of the larger size has been eliminated in calculating the ages. It is assumed that the total rate of production of He^3 is twice the observed rate for H^3. In Table 2*b* more approximate results are given using the constants listed in the table. Only the smaller iron meteorites are included because of probable high shielding of the larger ones. The K^{40}-A^{40} and He^4-U, Th, ages are given for comparison.

TABLE 2*a*—*H^3-He^3 ages of meteorites*

Meteorite and year of fall	H^3-He^3 age, my	H^3, dpm	Reference
	Stones		
Abee, achondrite, 1952	13	0.26 ± 0.04	3
Elenovka, chondrite, 1951	30	0.30 ± 0.03	8
Kunashak, chondrite, 1949	4	0.31 ± 0.03	8
Monte das Fortes, chondrite, 1950	40	0.30 ± 0.05	2
Norton County, bustite, 1948	230	0.26 ± 0.02	1
	Irons		
Norfolk, o. m., 1918	600	0.29	5
Para de Minas, o. m., 1934	1100	0.055	5

Some of the data are not very reliable; some diffusive loss of gases has probably occurred though it can hardly have been an importan amount (see Appendix), and the unknowi shielding for the objects listed in Table 2*b* undoubtedly exists. Yet it is evident that the cosmic-ray ages for the stones are markedly les than those for the irons, except Norton County and that they are much less than their K^{40}-A^4 and He^4-U, Th, ages as well. If the data are reliable, the stones and irons have had a very different history and have become separated fron each other in some way.

The following possible explanations for th data can be considered.

1. The shielding of the meteorites by thei larger size before entering the atmosphere cannot be an acceptable explanation for those liste in Table 2*a*, since the H^3 as well as He^3 wa measured.

2. A steady ablation of the surface of the ston meteorites by collisions with micrometeorites o by particle and light radiation from the su would constantly renew the surface. But about

ABBREVIATIONS IN TABLE 2*a* AND *b*

au.	aubrite	gg.	very coarse	o.	octahedrite
br.	bronzite	gr.	gray	pa.	pallasite
brec.	brecciated	h.	hexahedrite	poly.	polymict
bu.	bustite	ho.	howardite	sh.	shergottite
cr.	crystalline	hyp.	hypersthene	sph.	spherical
eu.	eukrite	int.	intermediate	v.	veined
g.	coarse	m.	medium	wh.	white

TABLE 2b—*Cosmic-ray ages of meteorites*

K^{40}-A^{40}, and He^4-U, Th, ages in aeons, if known, are given for comparison.

Stones: Assumed production ratio in cubic centimeters STP per gram and million years. He^3, 10^{-8}; Ne^{21}, 10^{-9}; A^{38}, 4×10^{-10}.

Chondrites and Classification	He^3, m y	Ne^{21} m y	A^{38}, m y	K^{40}-A^{40}, aeons	He^4-U, Th, aeons	Reference
kaba, wh. hyp.	90	...	...	3.6	3.4	13
eardsley, gr.	9	...	...	4.30	3.8	4
urböle, sph. hyp.	17	...	<39	4.32	4.2	4, 10
	15	17*	5	3.7	4.1	9
lenovka	30	50	11	3.9	4.1	9
	30	...	...	...	...	8
olbrook, cr. sph. hyp.	27	...	27	4.40	4.4	4
unashak, gr.	4	4	3	0.7	0.5	9
	4	...	...	...	...	8
ocs, v. wh. hyp.	...	...	37	...	...	10
onte das Fortes	40	...	...	...	...	2
ew Concord, v. int. hyp.	<2†	...	...	...	1.0	4
chansk, poly. brec. sph. br.	15	10	5	4.5	4.0	9
ervomaiskii, int. (black)	...	2	3	1.8	...	9
v. cr. enst. (gray)	6	8	5	0.65	0.63	9
ichardton, v. sph. br.	33	...	30	4.15	3.9	4
aratov, gr. sph. hyp.	30	45	15	3.7	4.0	9
. Michel, wh. hyp.	32	...	32	4.00	1.9	4
ovtnevyi, int. v.	40	60	19	3.0	4.3	9
chondrites						
bee	12	...	...	...	3.6	3
shopville, au.	31	...	...	...	...	4
oore Co., eu.	...	...	10	3.2	...	10
orton Co., au.	230	...	260	4.4	...	1, 10
uevo Laredo, ho.	4	4	...	3.1–3.6	0.45	14
advarninkai, eu.	...	19	20	1.0	...	9
asamonte, ho.	≥8‡	...	≥19‡	3.8	...	4
allowater, au.	>28§	...	...	...	...	4
ergotty, sh.	5	...	9	0.56	...	10
renham Township, pa. (olivine)	>105‖	...	<82	...	...	4

Irons: Assumed productions of He^3, 5×10^{-9} cc/g and million years.

	He^3, m y	Ne^{21} m y	A^{38}, m y	K^{40}-A^{40}, aeons	He^4-U, Th, aeons	Reference
arbo, o. m.	950					6
harcas, o. m.	520					5
enbury, o. m.	350					7
orse Creek, pseudo-octahedrite	4					11
ailac, pa.	150					7
orfolk, o. m.	1100					5
orfork, o. m.	600					5
ara de Minas, o. m.	1100					5
tts, o.	500					5
t. Ayliff, o. g.	1750					6
n Martin, h.	70					6
khote-Aline, o. gg.	...		500			12
amarugal, o. m.	1100					6
hunda, o. m.	1350					6
ropilla, h.	85					7

*It is assumed that this Ne^{21} age is the same as e He^3 age for Bjurböle as a means of establishing relationship between the He^3 ages and the Ne^{21} es.

†This is a lower limit. No He^3 was detected.

‡This meteorite may have been very large, since a large cloud of dust was observed; hence, screening may have been important.

§The extraction of He^3 was incomplete.

‖This was a large meteorite, and probably many fragments were not found; hence, screening may have been important.

Note: References of Table 2b are at the top of the next page.

References (for Table 2*b*)

1. F. Begemann, J. Geiss, and D. C. Hess, *Phys. Rev., 107,* 540, 1951.
2. J. Geiss, H. Oeschger, and P. Signer, *Helv. Phys. Acta, 31,* 322, 1958.
3. F. Begemann, P. Eberhardt, and D. C. Hess, *Z. Naturforsch.,* in press.
4. P. Eberhardt and D. C. Hess, *Astrophys. J.,* in press.
5. E. L. Fireman and D. Schwarzer, *Geochim. et Cosmochim. Acta, 11,* 252, 1957.
6. K. H. Ebert and H. Wanke, *Z. Naturforsch., 129,* 766, 1957.
7. W. Gentner and J. Zähringer, *Geochim. et Cosmochim. Acta, 11,* 60, 1957.
8. J. Geiss and H. Oeschger, private communication, 1959.
9. E. K. Gerling and L. K. Levskii, *Doklady Akad. Nauk SSSR, 110,* 750, 1956.
10. J. Geiss, *Chimia, 11,* 1957.
11. Based on datum from A. O. C. Nier.
12. E. L. Fireman, *Nature, 181,* 1613, 1958.
13. P. Reasbeck and K. Mayne, *Nature, 176,* 186, 1955.
14. J. H. Reynolds and J. I. Lipson, *Geochim. et Cosmochim. Acta, 12,* 330, 1957.

30-cm thickness is required for shielding, and this must be removed every few tens of millions of years. Thus, if the true age were some 5 x 10^8 years, about 25 times 30-cm thicknesses must have been removed, or 7.5 meters. One cannot expect that all stone meteorites started out some 15 meters in diameter and were worn down to 1-meter diameter before they *dared* to collide with the earth. Stone meteorites do not appear to be remnants of very much larger bodies than the largest ones collected.

3. Objects coming from the moon, removed from that body by collisions of iron meteorites coming from the asteroidal belt or by collisions of the residues of comet heads, should have orbits near the earth's orbit and should collide with the earth in shorter periods of time than objects coming from the asteroidal belt. In this case the stones came from the moon and the irons from the asteroidal belt. Possibly Norton County comes from the asteroidal belt, or it may be only an object that was lying near the surface of the moon and hence was irradiated before it was removed from the moon. The Horse Creek iron may be the rare object that by chance got to the earth quickly, or it may come from the moon. Its chemical composition is most unusual, for it contains some 2 per cent of elementary silicon [*Henderson,* 1958].

Öpik [1951] has calculated the times of arrival of objects on the planets as they depend on the orbital constants of the meteorites. Part of his data are given in Table 3. One notes that objects moving in orbits near that of the earth should have lifetimes equal to those observed for the stone meteorites, and that those crossing the orbits of Mars and earth should have lifetimes approximately the same as those of the iron meteorites. If Öpik's calculations and the data of Table 2 are both correct, the stone meteorites should not arrive from the asteroidal belt in the times observed.

But the stone meteorites must be removed from the moon by some process such as collisions of iron meteorites or comet heads with the moon's surface. This is not likely to be a very efficient process. Calculations bearing on the problem are very difficult. It should be noted that the Siberian meteorite of 1908 was the largest observed fall, and no meteorite fragments were observed. Also, the Carolina Bays are ascribed by many geologists to the fall of a shower of more than 140,000 objects. These craters are sometimes several kilometers in size, and again no meteorite fragments have ever been found [*Huchinson,* 1957]. It seems probable that such collisions are cometary. If a group of low-density objects, say of density 10^{-1} or 10^{-3}, fell on the earth, the atmosphere would cushion them to a great extent; on the moon where there is no atmosphere, there would be no cushioning effect. Also, craters produced on

Table 3—*Collisions*

	With the Earth			
a*	1.0	1.0	1.0	1.0
sin i	0.02	0.05	0.10	0.20
e	0.02	0.05	0.10	0.20
T, 10^6 yr	0.45	2.8	16.5	90

	With Mars			With Earth and Mars		
a	1.524	1.524	1.524	1.524	1.25	1.25
sin i	0.02	0.05	0.10	0.20	0.20	0.40
e	0.02	0.05	0.10	0.20	0.40	0.40
T, 10^6 yr	300	420	850	1950	250	410

*a, i, and e are the semimajor axis, angle between the orbit and the plane of the eliptic, and the eccentricity, respectively, of the meteoritic orbit.

e earth may be quite different from those
roduced by meteorites, e.g., oval and shallow
stead of circular and deep.[3]

Krinov discards as unreliable all meteorite
bits calculated up to the present except those
r five stones and one iron. One of the five
ones is Pesyanoe, which contains large amounts
primordial gases (see Table 4). Yet this
one's orbit as calculated requires that it pass
ithin 0.25 astronomical unit from the sun; at
is distance its temperature would be about
00°C, and it would surely loose its gases, which
has not done. The calculations of meteorite
bits from the observed data of fall are un-
liable.

4. It is possible that collisions in the aster-
dal belt produce stone meteorites of about the
roper size and that subsequent collisions de-
roy them, so that they have lifetimes of only
me tens of millions of years. A few of them
ter orbits such that they can collide with the
rth. Hence, the cosmic-ray lifetimes observed
e the life times of these objects as determined
collisions.[4] This is a possible explanation of
e facts, though it calls for a very special set
circumstances in the collisional processes. The
oss-sectional area of Mars is 10^{13} m^2. Taking
e cross section of a meteorite as 1 m^2 and us-
g an appropriate value for the collisional prob-
ility with Mars from Öpik's calculations as
5 x 10^8 per year, we find that for collisions to
cur in 2 x 10^7 years some 10^{20} grams of me-
orites must be present in places appropriate
r collision and the material must be replenished
ery 20 million years. In 4.5 aeons the total
aterial would be 2 x 10^{22} grains, which is still
ly a small fraction of the total asteroidal ma-
rial now present in that region. The suggestion
es not help us to understand the structures of
ondritic meteorites. It requires that the irons
far more numerous relative to the stones than
eir corresponding abundance in the primary objects, but this is not a valid objection to the proposal. Such collisions with the irons may account for the turned-in lips of the holes in these objects, for this number of collisions means that all the irons have been bombarded over their entire surfaces by stone meteorites.

TABLE 4

Nuclide	Suess and Urey He^4 = 1000	Gerling and Levskii He^4 = 1000
He^4	1000	1000
He^3	...	≤0.278
Ne^{20}	2.51	3.39
Ne^{21}	0.0084	0.029
Ne^{22}	0.278	0.31
A^{36}	0.0409	0.254
A^{38}	0.0078	0.054

Both the lunar origin for stones and the postulate of destruction by collision lead to the conclusion that more of the metal phase than the silicate phase has been produced by collision processes in the asteroidal belt, a conclusion very similar to that of *Urey* [1956] to account for the varying compositions of the terrestrial planets.

Assuming that the lunar origin is correct, some interesting conclusions about the lunar surface follow. Only fairly strong objects could be accelerated from the moon's surface, and hence a surface layer of dust or gravelly material may be present. However, appreciable amounts of material of a chondritic type must exist near the surface of the moon. It is reasonable to suppose that it was produced by the great collisions that produced the maria and large craters. The substance of this material must have come partly from the colliding object and partly from the primitive surface of the moon. We have used these chondritic meteorites as a proper sample of solar elemental abundances, and it may be necessary to question such an assumption, as I have done here. Also, some parts of the moon's surface must supply the achondrites. Are these the mountainous parts, or are they part of the maria which may have been partly melted?

Gerling and Levskii [1956] have found large amounts of the inert gases helium, neon, and argon in the Pesyanoe meteorite. The neon and argon have the isotopic composition of terrestrial neon and argon, and the amounts of these gases

[3] The origin of tektites proposed by *Urey and*
hers [1957] is a suggestion of the same general
ture; this theory is the only nonmiraculous one
r the origin of these objects, in my opinion.
e recent discovery of tektites in Georgia very
ar the Carolina Bays is most interesting.

[4] This suggestion was made by Dr. Peter Eber-
rdt during discussions after the symposium in
ashington.

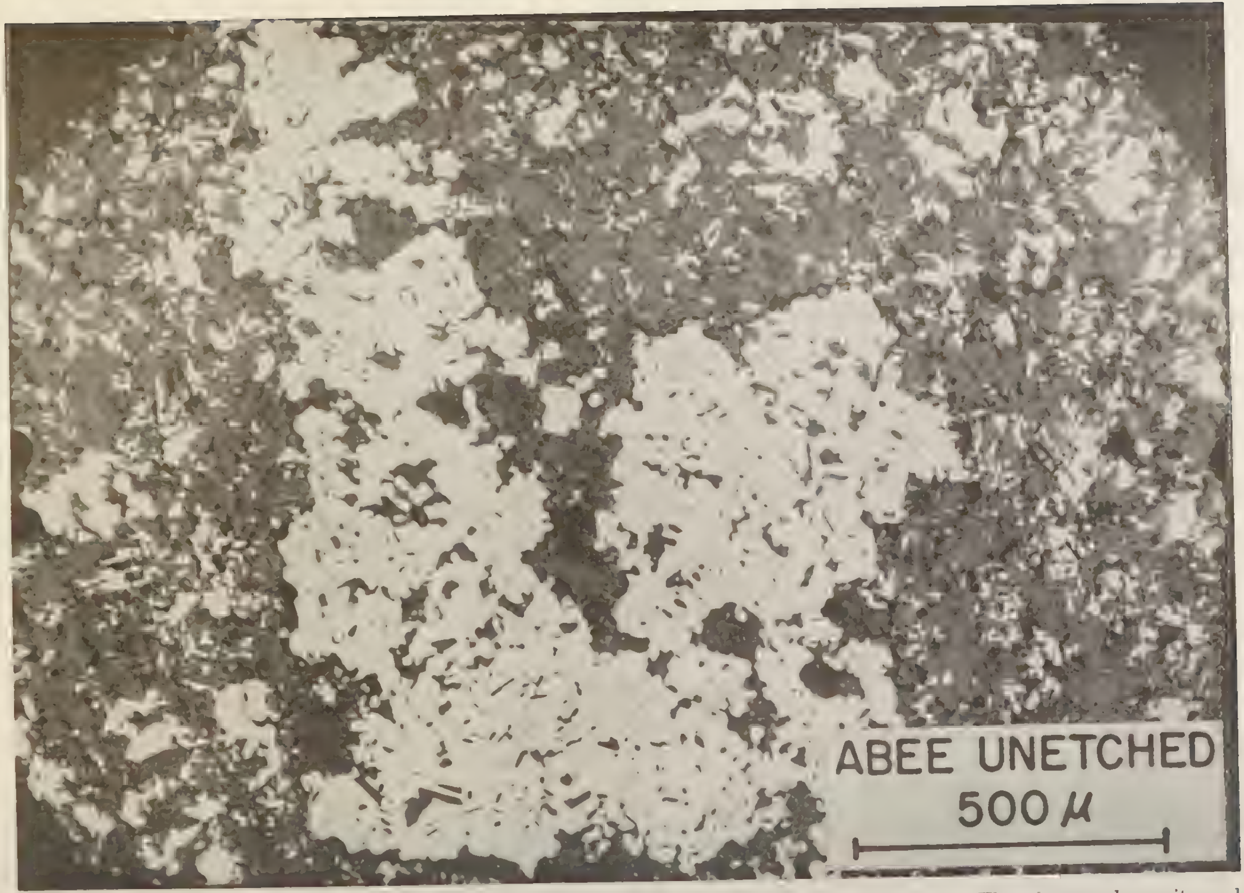


FIG. 9—Abee. This shows a distribution of metal particles quite different from that of the chondrites. Though some kamacite and taenite particles can be distinguished, most of the metal under etching shows a very different appearance. See Plate 8.

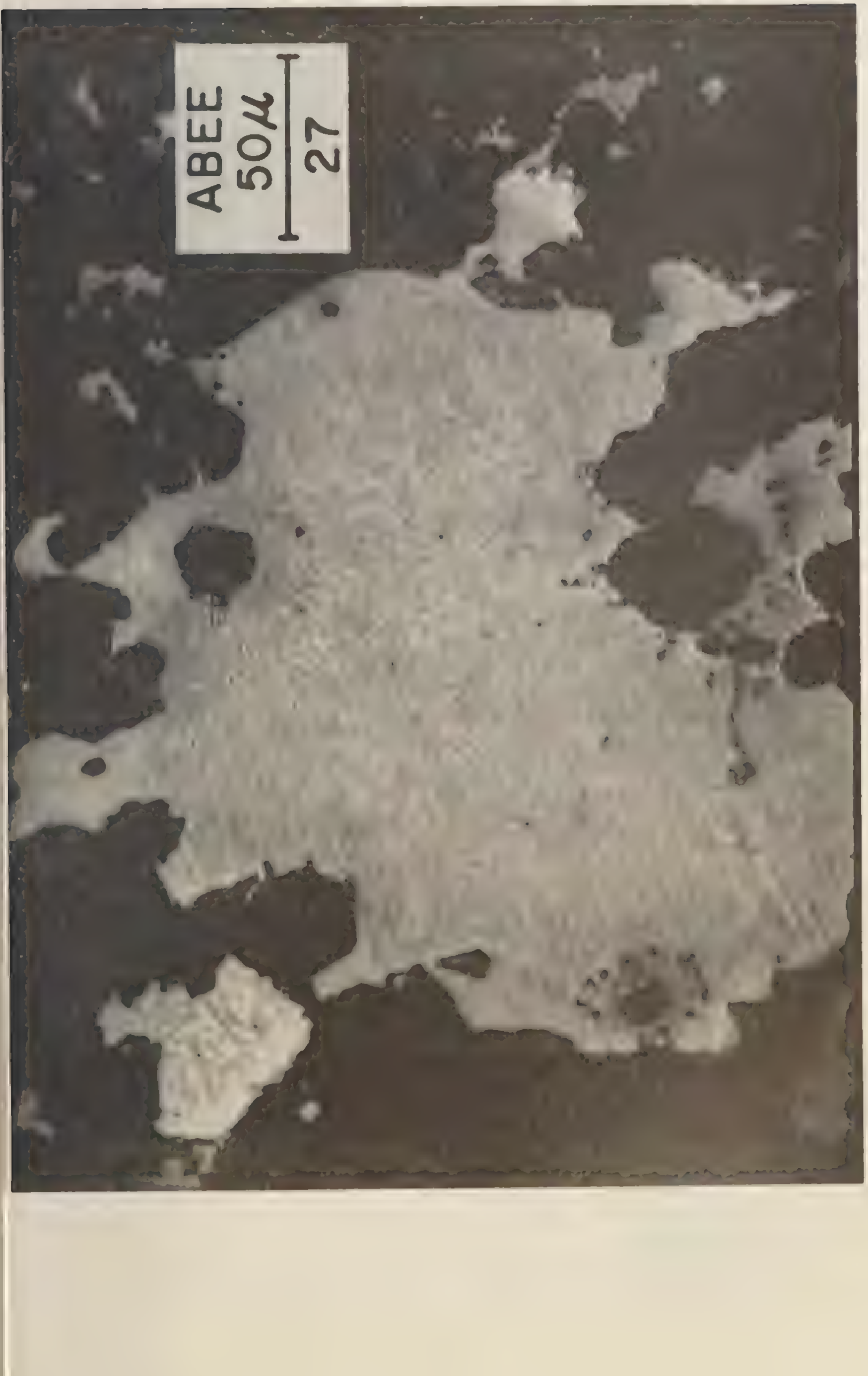


FIG. 10—Abee. The etch pattern is similar to that of some quenched steels.

and helium are 100 to 1000 times the amounts observed in other stone meteorites. Table 4 compares the relative amounts of these gases and their isotopes with the estimates of the Suess-Urey tables. Although agreement is not exact, it is sufficiently close to leave no doubt that these are primitive gases trapped in the meteorite long in the past. Gerling and Levskii estimate the partial pressures of helium, neon, and argon over a silicate melt required to produce the observed concentrations. If our interpretation is correct, the moon was subjected to 25 atmospheres pressure of cosmic gases at some time in the past, and if Pesyanoe comes from some other object than the moon, it was subjected to such gaseous pressures.

The Abee meteorite which fell in Canada some years ago has curious metallic structures quite dissimilar from those of the chondrites. Some kamacite and taenite can be found, but most of the metal gives an etch pattern similar to chilled steel, and silicate particles are suspended in the metal particles (Figs. 9 and 10). Mare Tranquillitatis looks like a very black lava flow. Might this black meteorite which probably formed from a melt possibly come from this mare?

Several experiments in a lunar explorations program will provide valuable information:

1. Seismic investigations may be able to detect or disprove the 'raisin bread' structure postulated for the primary objects including the moon.

2. If the moon contains enough water to make its content of iron agree with that of the chondritic meteorites, water should be an abundant constituent of the lunar surface rocks. Chemical analysis for water would give information about the elemental abundance of iron.

3. Chemical and physical analyses of the land and sea areas of the moon would decide whether the stone meteorites come from the lunar surface. Televised observations of the textures of broken surfaces of the lunar rocks would lead to positive identification of chondritic materials. Determinations of the radioactive elements in the lunar surface and the magnetic properties would give additional evidence.

It is hoped that such observations will be forthcoming during the immediate years ahead.

Appendix

Kunashak, Pervomaiskii, and Shergotty c Table 2*b* have very low K^{40}-A^{40} ages, and th question arises whether A^{40} and A^{38} have diffuse out of them. It is easy to show that the low A ages are not due to this effect.

The rate of increase in A^{40} is given by th equation

$$dN/dt = (0.124/1.124)\lambda N_K - kN$$

where N is the concentration of A^{40}, λ is the decay constant for K^{40}, 0.546 aeon^{-1}, k is a constant determining the rate of loss, and N_K is th amount of K^{40} present and is equal to $N_0 e^{-\lambda}$ where N_0 is the original concentration of K^{40} a the time when N was 0. Setting (0.124/1.124 $\lambda = \lambda'$ for convenience, and integrating,

$$N = (\lambda' N_0/k - \lambda)(e^{-\lambda t} - e^{-kt})$$

If we assume that the time since N was zero 4.5 aeons, $N_0 = N_P e^{4.5\lambda}$, where N_P is preser concentration of K^{40}, and then

$$N = (\lambda' N_p/k - \lambda)(1 - e^{-(k-\lambda)4.5})$$

at the present time. But N is such that the apparent age with no diffusive loss is 0.56 aeo for Shergotty, and hence for this meteorite

$$N = N_p(\lambda'/\lambda)(e^{+0.56t} - 1)$$

Equating these, we have

$$1/(k - \lambda)(1 - e^{-(k-\lambda)4.5})$$
$$= (1/\lambda)(e^{0.56\lambda} - 1$$

We can solve for k from this equation, since λ known, and we find $k = 2.17$ aeons^{-1}.

The equation for A^{38} whose concentration taken as N' is

$$dN'/dt = K - kN'$$

where K is the rate of generation of A^{38} by cosm rays. Integration gives

$$N'/K = (1/k)(1 - e^{-kt})$$

N'/K is our calculated age, and t is the true ag Assuming that k for A^{38} is the same as that fo A^{40} calculated above, we can calculate from th equation the value of N'/K. From Table 2 N'/K is 0.009 aeon. A little numerical calcul

shows that the apparent age is lower than
true age by about 1 per cent. If the time
e very large, the apparent age $N'/K = 1/k =$
aeon. It is evident that the A^{38} cosmic-ray
is not affected by diffusional loss.

he loss of radiogenic He^4 would follow the
ation

$$'/dt = 7\lambda_{235}N_{235} + 8\lambda_{238}N_{239} + 6\lambda_{232}N_{232} - kN$$

re the symbols refer to U^{235}, U^{238}, and Th^{232}.
integration of the equation gives

$$= \frac{7\lambda_{235}N_{235}{}^0}{k - \lambda_{235}}(e^{-\lambda 235t} - e^{-kt}) + \frac{8\lambda_{238}N_{238}{}^0}{k - \lambda_{238}}(e^{-\lambda 238t} - e^{-kt}) + \frac{6N_{232}\lambda_{232}}{k - \lambda_{232}}(e^{-\lambda 232t} - e^{-kt})$$

re the $N_{235}{}^0$ is the initial amount of U^{235} and
he present amount of U^{235} multiplied by
^{t}. The other definitions are obvious. The
ation when there is no loss, $k = 0$, is easily
ten. Some of the meteorites have He^4-U,
ages near 0.5 aeon. It is possible to solve for
ain, and its value is approximately 2 aeons^{-1}.
in the loss of He^3, assuming that its rate of
is the same as that of He^4, is found to be
igible.

he theory is approximate in assuming a
ple proportionability to the concentration
the rate of loss, since the actual diffusional
must follow a more complicated law de-
ding on crystal sizes, kinds of crystals, and
r factors. Since the laboratory observations
that He^3 and A^{38} are removed with greater difficulty than He^4 and A^{40}, the conclusion that diffusional losses of He^3 and A^{38} are negligible is safe.

Acknowledgments—I am indebted to Peter Everhardt for many illuminating discussions on the subject of this paper. This work was partially financed by AEC Contract AT(11–1)–34.

References

Aller, L. H., *Handbuch der Physik,* Vol. LI, 1958.

Gerling, E. K., and L. K. Levskii, *Doklady Akad. Nauk S.S.S.R, 110,* 750, 1956.

Henderson, E. P., Williams Bay conference on elemental abundances, September 1953, *Proc. U. S. Natl. Mus., 107,* 347 ff., 1958.

Huchinson, G. E., *A Treatise on Limnology,* John Wiley & Sons, New York, 1957; gives review of subject and references to the original literature.

Jerofejeff, M., and P. Latschinoff, *Verhl. russ. min. Ges., 2,* 24, 272–292, 1888.

Krinov, E. K., *Principles of Meteorites,* Chap. 2. Translation being prepared by H. S. Brown.

Öpik, E. J., *Contribs. Armagh Observatory,* no. 6,
Collision probabilities with planets and the distribution of interplanetary matter, p. 186, 1951.

Paneth, F. A., *Geochim. et Cosmochim. Acta,* V. *I,* 8–9, 1951.

Patterson, C., G. Tilton, and M. Inghram, *Science 121,* 69–75, 1955.

Perry, S. H., *U. S. Natl. Museum Bull. 184,* 1944.

Prior, G. T., and Max Hey, *Catalogue of Meteorites,* British Museum, 1953.

Schumacher, E., *Helv. Chim. Acta, 39,* 531, 538, 1956; *Z. Naturforsch., 11a,* 206, 1956.

Sorby, H. C., *Proc. Roy. Soc. London,* 1864.

Urey, H. C., *Astrophys. J., 124,* 625, 1956.

Urey, H. C., and H. Craig, *Geochim. et Cosmochim. Acta, 4,* 36–82, 1953.

Urey, H. C., H. Mele, and T. Mayeda, *Geochim. et Cosmochim. Acta, 13,* 1–4, 1957.

Wahl, W., *Geochim. et Cosmochim. Acta, 1,* 28, 1950.

Wasserburg, G. J., and R. J. Hayden, *Phys. Rev. 97,* 86, 1955.

Remarks on Mars and Venus

G. DE VAUCOULEURS

Harvard College Observatory
Cambridge, Massachusetts

ur investigations of Mars and Venus can be
sified into the categories of astronomy, as-
hysics, and geophysics.

he available astronomical data on these
ets are rather complete; astrophysical data
incomplete; and geophysical data are no
e than fragmentary. I shall review the astro-
ical data quickly, discuss the astrophysical
, at some length, and finally present what
we know of the geophysical information.

able 1 lists some astronomical elements of
two planets. Table 2 gives the elements of
globes of the two planets. Diameters, masses,
volumes are fairly well known. The mass of
us, being derived from planetary perturba-
in the absence of a natural satellite, is
efore known with relatively low accuracy.

he rotation period of Venus has not yet been
rmined. Estimates have varied between 24
s and 225 days, and we believe now that it
be of the order of some weeks. The inclina-
of the equator to the plane of the orbit
so uncertain.

regard to Mars, the only serious uncer-
y refers to the ellipticity of the globe. This
be determined in two ways: by measure-
ts on the flattening of the disk with a
ometer; and by observations on the motions
he Martian satellites. The satellite orbits
the dynamic ellipticity listed in the table,
the micrometer measurements give a flat-
ng two or three times greater. The reason
he discrepancy has not yet been found.

hat can space scientists contribute to the
ion of these problems? We could derive the
of Venus and the diameter of its globe
lacing an artificial satellite in orbit around
is, the orbital plane being in the ecliptic or
to it. If this satellite contained a radio
smitter of very stable frequency, it would
titute what the astronomer would call a single-line spectroscopic binary, and from the periodic variations of the received frequency due to the Doppler effect the elements of the orbit could be derived. The period and major axis of the orbit would give the mass of Venus with great accuracy. If the satellite further contained a radar system we could determine the distance to the solid surface, and by combining that with the Doppler data, we could derive the true diameter of the solid globe of Venus. As is well known, we cannot see the surface of Venus, which is covered by clouds.

Turning now to the astrophysical data, we see in Table 3 a summary of the composition of the atmospheres of the earth, Mars, and Venus.

For Mars, we know the total pressure, i.e., the total mass of the atmosphere per unit area, from various lines of evidence (photometry, polarimetry), and since carbon dioxide, the only gas that has been spectroscopically observed in the atmosphere of Mars, is present in only a very small amount, we must conclude that the bulk of the atmosphere is nitrogen. This constituent cannot be observed in the spectroscopic range accessible from the surface of the earth.

With regard to Venus, our only evidence comes from the carbon dioxide absorption bands, which indicate more than 1 km of carbon dioxide above some level fairly high in the atmosphere, probably above the cloud layer.

Table 4 is a survey of the available determinations of planetary temperatures. Shown first are the theoretical determinations of the average temperature (T_a) and maximum temperature (T_M) for a black body. The latter is the temperature of the subsolar point for a black body turning always the same face toward the sun and represents an absolute maximum for a planet without an atmosphere. The gray body is a more realistic case involving allowance for the albedo (reflectivity) of the planets. The

Table 1—*Orbital elements of Mars and Venus*

	Venus	Mars
Semimajor axis	0.723 au* = 108·2·10^6 km	1.524 au = 228·10^6 km
Sidereal revolution period	224.7 days = 0.615 yr	687 days = 1.881 yr
Synodic period	584 days	780 days
Mean orbital velocity	35.05 km/sec	24.15 km/sec
Eccentricity	0.0068	0.0933
Perihelion distance	0.718 au = 107·5·10^6 km	1.383 au = 206·10^6 km
Aphelion distance	0.728 au = 108·9·10^6 km	1.666 au = 248·10^6 km
Inclination to ecliptic	3°23′.6	1°51′
Longitude of ascending node	75°47′	48°47′
Longitude of perihelion	130°9′	334°13′
Longitude of vernal equinox†	?	87°

* Astronomical units.
† In northern hemisphere.

Table 2—*Elements of the globes of Mars and Venus*

	Venus		Mars	
Equatorial diameter	0.97 = 12,400 km	(*a*)	0.533 = 6800 km	(*b*)
Dynamical ellipticity	0.0	(*c*)	0.00522	(*d*)
Volume	0.91	(*e*)	0.150	(*e*)
Mass	0.815	(*f, g*)	0.107	(*f, h*
Mean density	4.95 g cm^{-3}		3.95 g cm^{-3}	
Surface gravity	850 cm/sec^2	(*i*)	373 cm/sec^2	(*j*)
Escape velocity	10.3 km/sec	(*k*)	5.0 km/sec	(*l*)
Rotation period	Weeks?	(*m*)	24h 37m 22.67s	(*n*)
Inclination of equator*	32°?	(*o*)	25.°0 ± 0.2°	
Satellites	0	(*p*)	2	(*q*)

* To orbital plane.

(*a*) refers to top of cloud layer; (*b*) refers to top of haze layer; (*c*) theoretical value; (*d*) value deri from the orbits of the Martian satellites, the optical value (0.013) is in excess of maximum theoret value; (*e*) earth = 1.0; (*f*) earth = 1.0; (*g*) from planetary perturbations; (*h*) well determined from sa lites; (*i*) derived from observed radius; (*j*) 375 cm/sec^2 at poles, 370 cm/sec^2 at equator; (*k*) at top of cl layer; (*l*) at surface; (*m*) see text; (*n*) direct sense; (*o*) see text; (*p*) satellites smaller than 25 km in ameter would not be observed; (*q*) Phobos and Deimos, estimated diameters 16 and 8 km, respectiv

Table 3—*Atmospheric composition of the terrestrial planets*

	Venus	Mars		Earth	
Gas	Thickness, m, STP	Thickness, m, STP	Volume, %	Thickness, m, STP	Volume, %
N_2	?	1650	93.8	6246	78.08
O_2	~0	<2	<0.1	1676	20.94
A	?	70?	4.0?	74	0.94
CO_2	>1000	40±	2.2±	2.2	0.03
H_2O	?	Very small		Variable	

Table 4—*Temperatures of the terrestrial planets, °K*

		Venus	Earth	Mars
3lack body	T_a	327	278	223
	T_M	464	394	318
3ray body*	T_a	229	246	217
	T_M	324	349	307
)bserved radiometric	T_m	230	200	160
	T_a	235	288	230
	T_M	240	350	300
	3.15 cm	580 ± 50	...	218 ± 50
	Rotational	285 ± 9	...	...

' Visual albedo: Venus = 0.76 (Kuiper) or 0.64 (Dajon); earth = 0.39; Mars = 0.15.

= minimum temperature.

= $394°/a^{1/2}$: black sphere facing sun at subsolar point (assuming solar constant = 2.00 cal/cm²/min); a = semimajor axis of orbit.

= $278°/a^{1/2} = T_M/\sqrt{2}$: black sphere in rapid rotation with negligible conductivity.

$_{ay} = T_{black}\,(1 - A)^{1/4}$, A being visible albedo.

nputed gray-body values for the earth agree ll with the observed values.

Also shown are the observed minimum (T_m), erage (T_a), and maximum (T_M) radiometric nperatures, measured in the infrared transssion window of our atmosphere between 8 d 15 microns.

Two other temperature determinations are ailable. One is the 3.1-cm microwave temperare measured at the Naval Research Laboray recently. For Mars it is roughly in agreeent with the observed radiometric temperature d also with the computed temperature. But Venus the discrepancy is large, a radiometric nperature of 580°K vs. a computed average 327°K or an absolute maximum of 464°K for black body.

Perhaps we must assume a very strong greenuse effect in the Venusian atmosphere, most the solar energy reaching the surface and little the planetary heat escaping back to space.

The rotational temperature is obtained from intensity distribution of the lines in the bon dioxide bands. This determination was de some years ago by Chamberlain and iper; it is reliable, but refers to an unknown ective level in the atmosphere of Venus.

We will now survey the models which fit se observations on Mars and Venus.

As a first approximation, Figure 1, taken m the *Space Handbook* compiled by the nd Corporation, shows the atmospheric denes of the three planets as functions of altitude. It is important to note the low value of the density gradient on Mars, associated with the small force of gravity. On Venus and the earth the gradients are similar.

In Figure 1 the surface pressure on Venus has been taken as 10 atm, on the assumption that the entire content of carbon dioxide present in fossilized form on earth has remained in

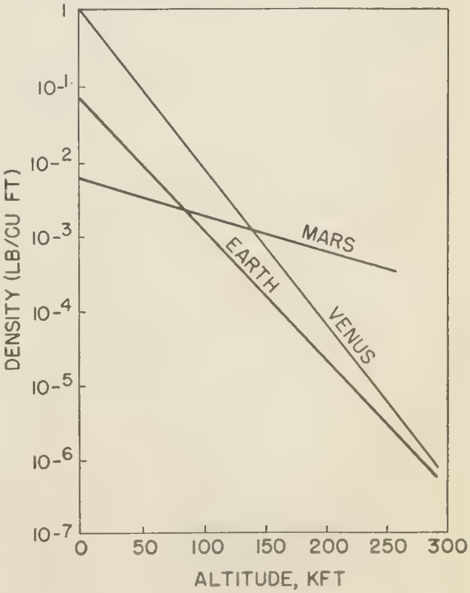


Fig. 1—Atmospheric densities of the terrestrial planets.

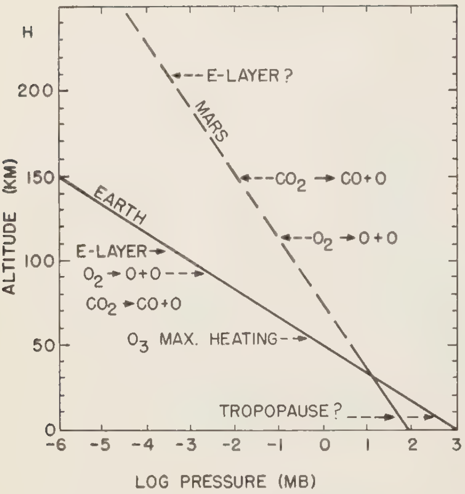


Fig. 2—Atmosphere of Mars, after Goody.

the gaseous state in the atmosphere of Venu (Dole).

The surface pressure of Venus can be est mated by other methods, with results close t this value.

Figure 2 represents the lower atmosphere (Mars. It is adapted from a graph by Good and also shows pressure vs. altitude for th earth, with several regions of special intere indicated.

We know that the surface pressure on Ma is approximately 85 millibars. We also kno the density gradient if we make plausible a sumptions about the temperature and structu of the atmosphere; the temperature gradient about 3.7°K/km for an atmosphere in conve tive equilibrium.

A model for the atmosphere of Venus is show in Figure 3. It assumes a surface temperatu of 580°K, on the basis of microwave measur

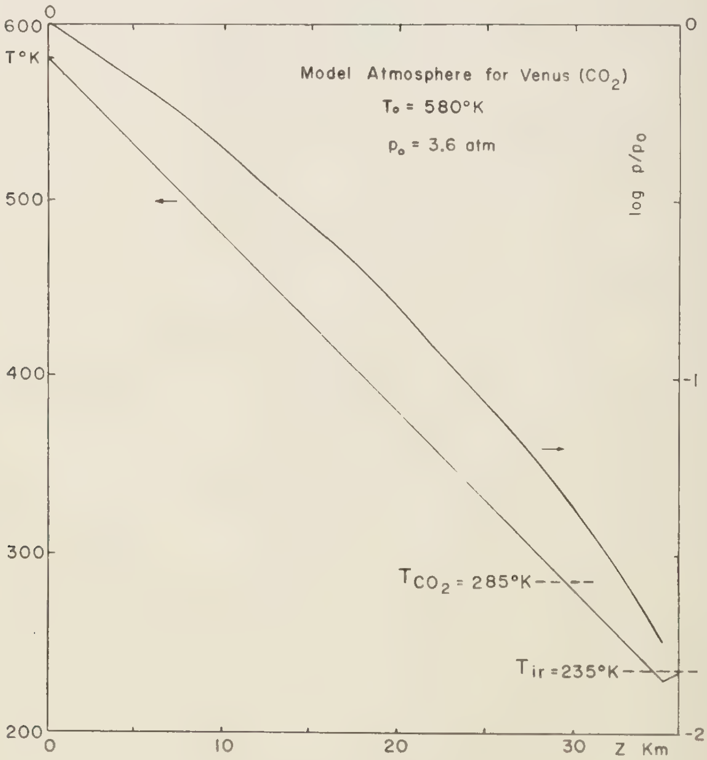


Fig. 3—Model atmosphere for Venus.

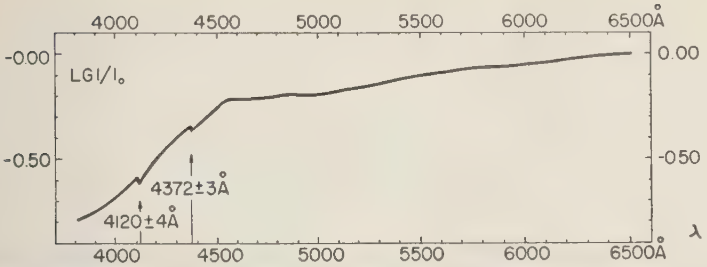


FIG. 4—Spectral reflectivity of Venus (after Kozyrev). The ordinate represents the log of the ratio I/I_0 of Venus to sun intensities with arbitrary zero point; abscissa is wavelength in angstrom units.

nents. The top of the cloud layer may be 235°K, according to the infrared radiometric lata. The altitude of this level, about 35 km, s determined on the assumption of a linear emperature gradient of 10°K/km, corresponding to convective equilibrium. The rotational emperature of carbon dioxide then refers to an ltitude of about 30 km. The relative pressure, p/p_0, follows from the temperature curve and he corresponding surface pressure inferred from he observed pressure of carbon dioxide above he 30-km level in 3.6 atm.

The surface of Venus is covered with clouds aving a yellow tint. The tint may be associated vith dust, since clouds of yellow surface dust re observed on Mars, and colored dust is also ery common on earth or of some white dust. f on the other hand we assume that the clouds re droplets of water as on earth, we may still xplain the yellowish color as a consequence of Rayleigh scattering in the gaseous atmosphere vith which the cloud layer is covered. We can ompute the amount of Rayleigh scattering eeded to change white into the observed color; ve find the equivalent of 1.5 atm or 3 miles of arbon dioxide above the cloud layer, i.e., above 5 km. This is obviously an upper limit to the ressure at this level.

Figure 4 demonstrates that, when we are oberving the reflection of light reflected by Venus, ur observations are not the result of pure scatering. The figure gives the log of the ratio of e spectral intensity of Venus referred to the pectral intensity of the sun. If this ratio is aken arbitrarily as unity at 6500 A, it is about .1 in the near ultraviolet. In the blue, yellow, nd red, Venus is only slightly redder than the n. Short of 4500 A there is an indication of definite absorption according to Kozyrev who measured these spectral intensities, including two absorption bands at 4120 and 4372 A. According to Urey and Brewer these bands may be due to CO^+.

The composition of the atmosphere of Venus can also be inferred from observations on the dark side. Occasionally, the dark side of Venus seems to be weakly illuminated, but these are visual observations, doubtful at best. However, some years ago Kozyrev took spectra of the dark side of Venus and identified a number of emission features, some of which appeared to be due N_2^+. This result was tentatively confirmed by Newkirk a year ago.

Figure 5 shows the spectrum of the dark side of Venus taken by Newkirk. Some of the emission regions appear to coincide with the bands observed by Kozyrev, but one is apparently new.

It must be noted that these observations are difficult because of the scattered light from the bright crescent, and should be repeated.

Turning to Mars, we remark first on the dark

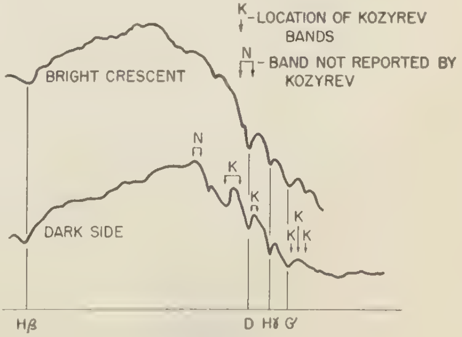


FIG. 5—Spectrum of the dark side of Venus (after Newkirk).

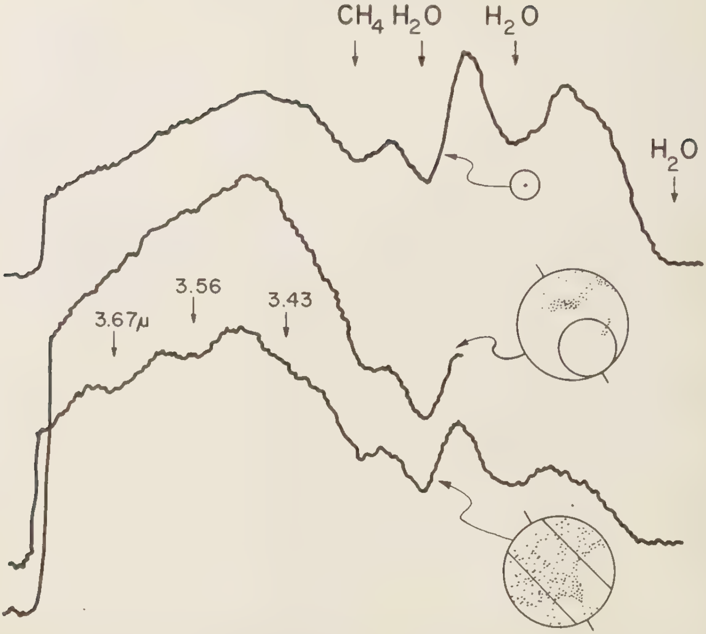


Fig. 6—Infrared spectrum of Mars (after Sinton). Proceeding from top to bottom, the curves show: the reflection spectrum of the sun with telluric absorption bands; the spectrum of a bright region of Mars; and the spectrum of a dark region of Mars, showing bands near 3.5 microns, which have been tentatively identified with bands observed in the reflection spectrum of terrestrial vegetation.

regions. These appear green in a refractor, because of the contrast effect or the secondary spectrum of the objective, and faintly bluish or neutral gray in reflectors. The true color is difficult to ascertain.

Dr. Kuiper has pointed out that the number of information elements in the area of the moon is about 10^8 with present terrestrial resolution. For Mars we have no more than 10^3 or 10^4 elements, indicating the limited amount of information obtainable from observations on the surface of this planet.

Figure 6 shows the most important discovery made on Mars during the past 30 years. It illustrates observations made by W. M. Sinton, of the Lowell Observatory, who worked with the 200-inch telescope at Palomar last October to observe the infrared spectrum. The three curves, from top to bottom, show the reflection spectrum of the sun, in which we see the absorption bands of methane and water vapor; the spectrum of a bright region of Mars, which is very much the same spectrum except for the change in slope produced by the color of Mars; the third tracing records three bands near 3.5 microns, which appear when the slit was set to the dark regions. These are not visible in the bright regions, indicating that the absorption occurs on the surface of the dark regions rather than in the atmosphere.

These bands are also observed in the reflection spectrum of terrestrial vegetation, and they are known there to be due to the vibration of the C-H bonds in organic molecules. It must be noted here that we are dealing with heavy molecules in which the mass of the molecule is very much larger than the vibrating bond, hence the wavelength is not the same as in a light molecule like methane.

Sinton's observation is very suggestive of the presence of organic molecules. In view of the earlier evidence on the seasonal variations and irregular changes in the dark regions of Mars, this remarkable result comes very close to offering the final proof of the existence of plant life on Mars.

Round-Table Discussion

Chairman: BRUNO ROSSI

Massachusetts Institute of Technology
Cambridge, Massachusetts

Participants: DINSMORE ALTER, A. G. WILSON, R. B. BALDWIN, W. M. SINTON, ROBERT DIETZ, THOMAS GOLD

Mr. Rossi: I should like to start by introcing the panel members. First, Dr. de Vauuleurs, of the Harvard Observatory; Dr. nton, of the Naval Observatory; Dr. Newell, ssistant Director of Space Sciences, NASA; r. Baldwin, from the University of Michigan d the author of the book *The Face of the oon;* Dr. Alter, Director of the Griffith Obrvatory of Los Angeles; Dr. Dietz, Marine eologist of the Naval Electronic Laboratory, n Diego; Dr. Wilson, of the Rand Corporan, former Director of the Lowell Observatory; d Dr. Gold, of the Harvard Observatory, and speakers in this session.

I will start by asking Dr. Sinton to say a v words about his most interesting work. It ms to me that the problem of the detection life on other planets is possibly the most exing and potentially the most important result space exploration. It is very important that start out by having that information from restrial observations.

Mr. Sinton: You all have seen the slides ich Dr. de Vaucouleurs showed during his narks. He mentioned the band at 3.67 microns. Originally I made spectra of terrestrial plants comparison with spectra of Mars. I did not lude this region in the investigation. When I n saw that Mars had some structure here, had to go back and look at the plants more refully, and at some different types of plants. r example, algae; I now find that algae have band at 3.67 microns.

You can find this band in filter paper, which a highly refined form of cellulose. It apparly is due to the presence of oxygen within anic molecules. The oxygen is attached to a bon atom which also has attached a hydrogen atom. The oxygen shifts the wavelength of the resonance to a longer wavelength, about 3.67 microns. It seems interesting to me that this correlation has come first from looking at the plants, then at Mars, and then at the plants again.

It seems very likely to me that the best explanation for the variation in intensity of the dark regions of Mars as mentioned this morning, and also for the observations of these bands, is that we do have plant life of some sort on Mars.

Mr. Rossi: What about the persistence of the patches after dust storms, and their regeneration?

Mr. Sinton: Yes, evidence has long been cited that the dark patches come back after having been covered with dust. During 1956, there was an intense dust storm which covered practically the whole surface of the planet, and yet these dark markings have come back.

I think it would seem that there must be some sort of life, something that has the power to shake off the dust, or to rejuvenate itself.

Mr. Gold: How long does it take for the dark patches to recover after the dust storms? Do they remain covered completely for some time, and then appear again?

Mr. Dietz: In 1956 a planet-wide dust storm cleared in about 6 to 10 days in certain of the dark patches.

Mr. Urey: It would appear that the plants of Mars have a capacity to dust themselves off.

Mr. Kuiper: I do not think it is a compelling argument, the one which was, I believe, first explicitly formulated by Öpik, that the dark areas have a regenerative power and that this proves the presence of organic matter.

If one flies over lava fields in the western part of the United States, one is impressed with the fact that these fields, some of them perhaps millions of years old, are still visible. And if one examines them on the ground, one notices that the surface is vitreous, smooth, glassy, there are many cracks, and the dust is blowing in the cracks. A dust storm can first cover the lava field, then the next wind can sweep the dust away into the cracks or sand. So I am inclined to think that regenerative power of these dark areas of mass is not proof that organic matter is present.

Mr. Jastrow: I should like to ask Professor Kuiper how many extinct volcanoes he believes there are on the lunar surface. What is their density?

Mr. Kuiper: The number of extinct volcanoes of the type I pointed out on the slides is of the order of 20. But there are quite a large number of even larger structures, about 4 times as large in diameter, with still lower slopes of the order of 1°. They look like blisters. Usually these objects are just little tips of spheres that seem to stick out of roughly horizontal terrain. They usually have a caldera right at the top of the order of half a mile across.

So there are two kinds of volcanoes, at least two types of shapes: those with slopes of the order of 1°, and those with slopes of the order of 5°. Then there are a great many very small structures, and the large domes, as in the maria. I think the number of features of this type on the moon runs into several hundred, if one takes all the various classes combined.

Mr. Jastrow: Have you a comment on the report by Kozyrev of the active lunar volcano?

Mr. Kuiper: This is a most difficult question to answer. I have been very much puzzled by this report, and I have had a lot of correspondence with my Russian colleagues on the matter. I don't believe that we have the information at the moment to make a critical statement on the correctness of the observation.

I have asked my Russian colleagues to look at the original spectrum and be sure that it is genuine. So far I have not been able to get anyone to state, 'I have looked at this spectrum and I am sure that it is genuine.'

Mr. Dietz: I should like to make a few comments about Professor Kuiper's excellent paper.

First, in regard to Alphonsus and Kozyrev, I think it is interesting to note that this crater is an old one. It antedates the Imbrium impact, and therefore one would think it would not be active.

A minor point with regard to the lunar surface: The words 'highland' and 'land area' were used. I should like to encourage the use of the word 'terra,' which contrasts with 'mare,' to indicate the so-called land areas.

As a geologist, I have a rather renegade opinion, which is that there is no volcanism on the moon in the classical geological sense. I have looked at it for some time and find no entirely convincing evidence of volcanism. I ascribe the features entirely to impact phenomena.

In connection with impact, you have a certain amount of melting, both superficial and deep melting beneath the surface. There are therefore secondary volcanic and volcanic-like effects which may or may not be called volcanic, but are not volcanic in the classical geological sense.

The moon of course is a geophysically small body compared with the earth. It has a very strong crust, and the pressure at the center is equivalent to that at 100 miles in the earth. For these and other reasons, the lack of deep holes in the crust suggests to me that the volcanism is most unlikely.

There are many central cones in lunar craters. I was pleased that Professor Kuiper did not seem to believe that there were any craters in the top. This is a very critical point; in the past there has been a good deal of discussion about craters in the tops of these central cones. I believe that in the long run the central cones will be explained on the basis of hypervelocity impacts, that is, impacts striking the moon in excess of the velocity of speed in lunar rock or in excess of about 6 km/sec.

Professor Kuiper also distinguished different types of craters on the moon. It is true that there is a great variety, but I think that they can be integrated as a family. You find a gradation in characteristics from small to large. I think this gradation depends upon the size of the impacting ballide plus the velocity of its impact.

In line with what Professor Urey said about doing research economically with meteorites, I believe we can learn much about lunar craters studying the earth. There are many lunar aters on the earth, not recent ones such as ie Meteor Crater, but the structures referred as cryptovolcanic, which I prefer to call yptoexplosion structures. Some of these reorded meteorite impacts are marked by a very nusual structure known as a shatter percuson cone in the middle. They are highly denged circular areas of rock. A locality for is type is in the Steinhein Basin of Germany, here the physiographic form is still preserved. here are at least three in the United States: at entland Quarry, Indiana; at Crooked Creek, issouri; and Wells Creek Basin, Tennessee. hese merit careful study.

Mr. Gold: With regard to volcanoes, I have ied to estimate the consequence of an average-zed active volcano on the moon in terms of sible effects. The activity of terrestrial volnoes would produce a number of plainly disrnible effects, not merely borderline observaons.

For example, the amount of gas that comes t of a terrestrial volcano when it is active enormous compared with the total gas content at we can allow in the lunar atmosphere.

It is hard to escape the conclusion that an dinary terrestrial volcano would be plainly sible in many ways. So that this, whatever was that happened, is certainly a very attenued version of a volcano, and possibly just e escape of cold gas.

About the question of craters in the central nes, which was mentioned, I should say that high-velocity impact is very likely to lead to small hole remaining right in the middle of e area that was blasted out, so that, even if ere are central holes in the central domes of aters, they would not make any difficulty for e impact explanation.

The phenomenon arises in this way: a part, obably quite a small part, of a high-velocity ject that comes in remains and survives the neral explosion that is produced; when it has en decelerated by the impact to below the eed of sound in the lunar material, it is able penetrate lunar material, at perhaps riflellet speeds, for a distance, leaving a hole, as think may have occurred in Meteor Crater, izona, where there is some indication that terial is buried deep below.

About the smoothness of the surface: it is most important to consider the radar evidence from the moon that shows the surface features to be extremely smooth with angles of more than 15° occurring only over a small fraction of the surface. On a scale of a kilometer the angles that we see are commonly only a few degrees, and occasionally up to 15°. On the earth you would find that if on a scale of a kilometer you see an angle of 15°, as in mountainous terrain, when you look at a scale of 10 cm the surface is most likely extremely rough. On the moon the surface is certainly a great deal smoother than that, according to radar data. The radar gives us the roughness on a scale of 1 wavelength or 10 cm. And that roughness is remarkably little, more like Sahara dunes than like Rocky Mountain terrain.

I proposed some years ago that an erosion process was responsible for the smoothness. That some kind of erosion has occurred is, I think, undeniable. We see the old craters very badly worn. Even if they are big in diameter they are very much smaller in height. The question is, what is the eroison process?

Of course, since there is no atmosphere and no water and so on, we have to think in quite different terms from terrestrial erosion. We have to think of the major influences on the lunar surface. The problem is, can we see any way in which material would be broken up into small particles, dust particles, or would perhaps be in small particles, and then can we see any way in which these small particles might be transported over large areas?

It is easy to suggest various reasons why there should be much dust on the surface of the moon. But the motion is a problem. I have suggested that electrostatic forces may be very significant for agitating the dust, electrostatic forces arising from either high-energy-particle bombardment or high-energy photons.

We have tried the high-energy-particle bombardment in the laboratory. A graduate student at MIT, Mr. Horwitz, set up an experiment in which he exposed dust, i.e., rock powders, to an electron bombardment of a few microamperes per square centimeter at electron energies between 70 ev and 1 kev; he found that individual particles of the powder jump around and make sizable individual jumps, and that the mean rate of movement is large. These particles are of the order of 25 microns across, and they are frequently made to jump distances of several millimeters at a time. If one tries to make some extrapolation of these results to the particle bombardment that we think the moon is getting from the solar wind, and from the solar outbursts that make magnetic storms, this rate of transportation would seem to be very high, much higher than is required for the supposition that the dust has filled the big areas on the moon.

Mr. Rossi: The solar wind is neutral as a whole. How do you separate positive and negative particles that produce local electrostatic charges?

Mr. Gold: It will not matter on a scale that is small compared with the Debye length in the gas whether the material is brought in neutral, provided that the secondary emission coefficient with the existing proton bombardment is substantially different from that for the existing electron bombardment. If, for example, solar wind consisted of 5-kv protons and 100-volt electrons, as is a very likely situation, then the 100-volt electrons will just be completely inoperative on a scale small compared with the Debye length.

Mr. Alter: I should like to say three things about Kozyrev. First, Kozyrev had been observing the crater Alphonsus systematically for some time, initially because of observations that I published reporting a haze in the same crater in 1956.

Second, at the end of a half hour, he said everything was exactly as it had been before. I am certain that there was no eruption in the ordinary sense. The word 'eruption' is technically correct, but it gives an absolutely wrong impression. I am convinced that there was a slight residual outgassing from a crater that may not have erupted in the last 2 billion years.

I do not know whether the Swan spectrum is there or not. I examined the photograph that he sent me, and we have had tracings made of it in the microphotometer, but I am unwilling to express an opinion. I am certain, however, that three emission broad bands are shown on it.

Mr. Baldwin: I have one comment I should like to make about the uniqueness of Mare Imbrium.

There are six similar easily visible circular structures on the face of the moon. Around each one of them there are certain radial markings. In none are they as clear and as well defined as they are around Mare Imbrium.

Around each one of them there are circular raised arcs that may be considered a frozen shock wave. The shock of impact sent these circular ridges of rock racing out, and they froze in position. The feature is clearly evident at the Alpine mountains around Mare Nectaris; it is very definite at Mare Crisium; less definite at Mare Humorum. The interpretation at Mare Imbrium is still somewhat in question. There are two structures around that, each one of which could be considered that shock wave. One is the scarp of the Apennines, and the other is a much broader feature farther out.

Regarding the Alpine valley, Dr. Kuiper referred to that as a graben, and it may very wel be. But radiating out from Mare Imbrium, and to a lesser extent the other circular mare, there are certain valleys which are not graben. The were produced by the blasting away of the surface by low-velocity projectiles, about 1 mile/sec ejected from the explosion itself. They are broad near Mare Imbrium, and become progressivel narrower farther out. They have a very decide preference for the uplands; that is, they ma go through a crater wall, not touch the insid of the crater, and appear on the far side. Th higher spots are definitely the ones affected b these markings.

Radial to Mare Nectaris are a small numbe of very large valleys. Their nature is not unde stood. Dr. Alter has published papers indicati that they probably are graben-like.

Mr. Urey: I should just like to say brief that I greatly admired the beautiful pictur Professor Kuiper has taken of the moon. The are certainly the best slides I have been priv leged to see.

On the other hand, I do not think that o can assume that one accretes a moon out average material, then partially melts it, a has the original accreted material still floati on that matter.

Water is the only common substance whc solid floats in it. There are a few others—ve few. I believe silicate solids do not float on th liquids. I therefore have serious objections model of the moon that makes use of this ypothesis.

Mr. Kuiper: I would say in answer to Dr. rey that it is my understanding that silicates ry rarely completely melt. I had been assured, Dr. Rubey, for instance, that in a mass of icate no more than 20 per cent of the mass ll be molten at any one time.

I believe that there is no contradiction in my sumption that the period of mare formation incides with the period of mare melting in e interior of the moon. That does not mean at the entire moon was molten at any one ne.

Mr. Baldwin: I agree that the period of maxium melting was relatively close to the time of e formation of the great circular maria. But you look at the pictures, in every single one the six there are craters within the mare ich are younger than the great dry crater at was originally formed, and older than the a flows.

To my mind the lava very definitely came er than the formation of the mare, and was t due to the energies released at the impact.

Mr. Urey: I should like to say that I would very glad to see something in the literature at could be studied carefully to clear up the int on the floating of solid silicates in their uids.

Mr. Jastrow: With reference to the bearing *in situ* lunar observations on these problems, ew things will come almost immediately from most primitive stage of lunar exploration. st, we will measure the extent of the lunar nosphere in one of our earliest projects, and t measurement will be a most interesting one. do not expect very much of an atmosphere, t how much is there depends indirectly on the gree of volcanic activity, for example.

Second, we will at the earliest possible date e what might be called snapshots from an tude of some hundreds of kilometers above surface, which will give us a resolution in detail of surface features substantially ater than that obtainable with the best tertrial telescopes.

Third, we will monitor the surface radioivity of the lunar crust, and we will probably that several times and with several different es of detectors. The results of that measurement will settle some of the present controversy on the composition of the lunar surface.

Mr. Rossi: Perhaps the question may be put: if you can walk around on the moon's surface, would you know whether the craters were of explosive or volcanic origin, for example? Surely the answers to most of these questions will be immediately obvious. We would settle how much was eroded dust, what was lava, and how much of that was volcanic.

Mr. Jastrow: That is the way in which one proceeds always. The questions that disturb those of us who are very interested in the moon at present will be immediately answered by the early unmanned stages of lunar exploration. For example, with respect to the magnetic field, one can make a guess about the magnitude of a lunar magnetic field, and the best guess probably is that there is no appreciable lunar magnetic field, for, it is believed, the earth's magnetic field is produced by currents circulating in its core, and the moon is smaller and generally considered to be colder and not to have a liquid core.

If there is an appreciable magnetic field, that will be a problem to deal with later. But we will measure the field, and if there is none, that result will support our present ideas on the lunar interior.

Then, again, at what might be called a late stage in the early phase of the lunar exploration program, we hope to be able to install devices for crude remote-controlled spectroscopic analysis of lunar samples by X-ray fluorescence techniques.

These likewise will confirm or deny the validity of such speculations as Professor Urey made this morning on the identity of the chrondritic meteorites with lunar material.

On the general question of why we wish to explore the moon, I believe—and I have been stimulated in this by contact with Dr. Urey and others—that the moon is perhaps the most interesting easily accessible object that we have in the solar system. It has a surface that is relatively unmarred by erosion, and on which the history of many aeons in the past has been written.

To a biologist, Mars and Venus are bodies of particular interest, but if the origin of the solar system is your concern the moon is the more interesting to you, and the major problem is to get close enough to read the history written on its surface.

Mr. Urey: Mr. Jastrow has stated the proposition very well, and I do not see that I need to make any comments except to say that the moon is certainly the least-changed object that is accessible now to us for investigation. Hence we can expect to get more of the record of the past from the moon than from anywhere else.

You will see, on the basis of the arguments about the moon presented today, that only one thing is certain: that people do not agree about it. Some see lava on the moon, some see pumice, some see dust, and some see chrondritic meteorite material.

I think, therefore, that if we are going to settle these questions it will be necessary to land instruments on the moon that can send signals back giving some critical data about its surface.

My own view is that the moon probably contains some lava flows. I have been able to see them. Mare Tranquillitis is a marked example. There is reason for believing that there is dust on the moon, as Dr. Gold has maintained. And I think there is really good reason for believing that there is some kind of stone meteorite material on the moon, or else we must explain the data that I presented today in some other way now unknown, which is of course always possible.

I am only elaborating what Dr. Jastrow said. I believe that, from the standpoint of the origin of the solar system, investigation of the moon is perhaps more important than investigation of Mars and Venus.

Mars, if it has life, would furnish information of remarkable interest, namely, that life could evolve somewhere else than on earth. But from the standpoint of the origin of the solar system, the moon is by far a more interesting object, because of its proximity and accessibility to exploration, and because of the relatively slow rate of change of its surface features.

Probably Mars once had oceans, implying erosion of some kind, and the covering of the whole surface with all the debris of erosion such as we see on earth. It seems reasonable to believe that Venus at one time had water on it, even though all the hydrogen may now have escaped. In this case the whole surface again has been covered with erosion materials. On the earth it has been exceedingly difficult to find out anything about the interior because of covering of the earth with the materials that result from erosion. On the moon we may have had such erosion as Dr. Gold has suggested. I think, however, that there are very good reasons to believe that the dominant features of the moon are not due to this effect.

Mr. Rossi: The erosion by dust formation has certainly not removed structures that are very much older than any that have occurred on the earth. The time scale for these erosive processes is extremely long compared with terrestrial erosion. Some of the structures we see on the moon are presumably a few billion years old, probably 3 billion years, or something like that. And we have nothing on the earth with that dignity.

Mr. Urey: May I make a remark on the Kozyrev problem? I have done a little spectroscopy in my life, although I cannot claim to be an authority. But I do think that the pictures that have been sent by Kozyrev to the United States either represent a real spectrum taken on the moon or an exceedingly elaborate piece of fraud, and I should like to say that it would be well not to accuse a scientist of such a serious charge without more information than seems to be available.

Mr. Alter: I should also like to return to the Alphonsus question. Related observations have been reported since. Wilkins in England observed a dark reddish spot on November 19. G. A. Hole independently made the same observation on the same night, using a 24-inch telescope at Brighton, and photographed it.

These men knew about Kozyrev's observations, but Dr. Poppendiek in San Diego is an amateur astronomer as well as a professional physicist who has been observing the moon for a good many years, and who had not heard of Kozyrev's observation. He looked at the moon on that same night and saw a white cloud over the central mountain of Alphonsus.

The information then that we have obtained from independent sources other than Kozyrev seems to me to be entirely convincing.

Rocket Astronomy

HERBERT FRIEDMAN

U. S. Naval Research Laboratory
Washington, D. C.

Rocket Astronomy—Whether he works from n observatory on the ground, or receives his ata from a rocket, an astronomer relies priarily on two tools, the telescope and the specrograph. The problems of conducting astroomical research from rockets, however, are uite different from the problems of classical stronomy.

To understand why we want to do rocket stronomy under inevitably adverse conditions, onsider the penetration of ultraviolet and X adiation into the earth's atmosphere as shown Figure 1. This figure is plotted for the level f 1/*e* penetration, which is the region of greatst interaction between the radiation and the tmosphere.

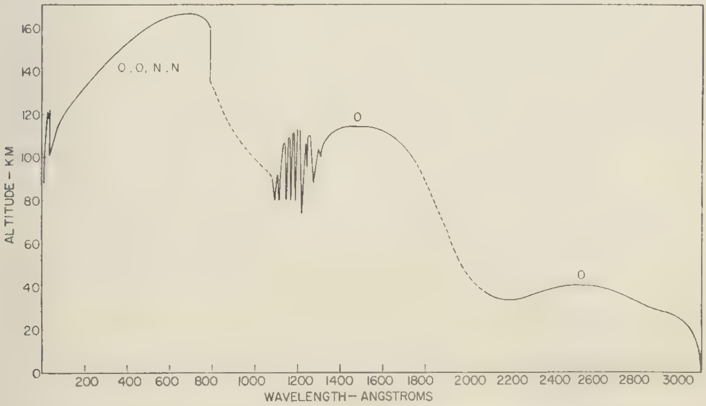
Immediately below 3000 angstrom units zone becomes an opaque barrier. At waveengths shorter than 2000 A ordinary molecular xygen absorbs so strongly that we must get ven higher up before we can see through the tmospheric gas. Between 1300 and 1000 A here are narrow windows in the atmospheric curtain. The one that we are most interested in is the window that falls at 1216 A. This window coincides almost perfectly with the wavelength of Lyman-α line of hydrogen.

Below 900 A, the depth of penetration of the radiation increases rather quickly to 160 km with the sun overhead. Absorption is due to ionization of all the atmospheric constituents.

The shortest wavelengths illustrated are in the X-ray region. As the X rays get harder they become more penetrating. At short enough wavelengths X-ray emissions penetrate to balloon altitudes.

We see that it is not necessary to get into outer space to do rocket astronomy. At 300 km we are well above almost all the absorbing atmosphere. In fact, as we go higher we run into trouble from the intense fluxes of the Van Allen radiation belt; it becomes a serious problem to shield detectors from the effects of bombardment by high-energy electrons and protons.

From the very beginning of the rocket program in this country, rocket astronomy was



IG. 1—The graph shows the altitude at which a fraction e^{-1} remains of the radiation incident on the atmosphere.

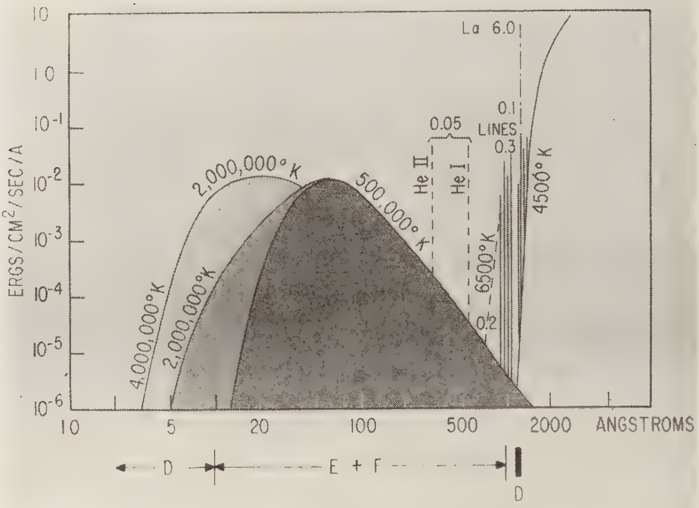


FIG. 2—Early results on the solar spectrum.

given an important place, and success came quickly. In 1946, Dr. Tousey and his co-workers at the Naval Research Laboratory obtained the first extension of the spectrum into the ultraviolet. They succeeded in pushing the spectrum down to about 2400 A.

From that time on there has been continued effort to extend the spectrum to shorter wavelengths. It appeared very early that simply allowing the spectrograph to spin and yaw with the rocket would reduce the exposure time to the point where it would be impossible to photograph the much weaker solar intensities in the the shorter-wavelength range. It therefore became necessary to develop pointing controls which could compensate for the gyrations of the rocket and maintain the spectrograph pointed at the sun.

The early experiences with these devices were discouraging. They failed for mechanical reasons, for poor rocket performance, and so on. It seemed that it would be well worth while to develop non-dispersive methods of looking at the spectrum and to acquire a broad picture of the energy distribution in the solar spectrum even if we had to settle for wavelength regions 100 to 200 A wide.

Figure 2 is a rough map of the solar spectrum as it was pieced out of data received from non-dispersive detectors plus some of the information obtained from the early spectrographs. This figure is a convenient starting point, and I will proceed from here to some of the most recent results, which will show what gratifying progress has been made.

It was quickly learned that the effective temperature of the sun in the continuum decreases rapidly into the ultraviolet and may fall to as low as 4000 to 4500°K in the neigborhood of 1200 A. Below about 1700 A the spectrum shows bright emission lines on an almost negligible continuum. The most interesting line in this range of the spectrum is the Lyman-α line of hydrogen, which is by far the brightest single emission wavelength in the solar spectrum. In the last few years most of our measurements have indicated that its flux is about 6 ergs/cm^2/sec at the outer atmosphere of the earth.

For a long time we had very little information on the spectrum below Lyman α. We theorized that the He I and He II resonance lines would be important, but could only estimate their contribution on the basis of what was required to produce the known ionization of the F region of the ionosphere with rather broad assumptions about recombination coefficients. On that basis

A total energy of 0.05 erg/cm²/sec was assigned to the two lines.

In the soft X-ray region it is again possible to use rather simple techniques. X-ray filters, of such elements as carbon, aluminum, nickel, and beryllium, form the windows of photon counters or ionization chambers. Combined with gases, which show increasing absorption toward longer wavelengths, they give us bandwidths of the order of 10 or 20 A.

With these detectors parts of the X-ray spectrum have been pieced together over the course of a solar cycle. The X-ray emission appears to have a broad base, centered at about 50 A, which follows the same variations over a solar cycle as the red line of Fe X whose ionization potential is 255 electron volts. This is an iron atom which has lost 9 electrons at the high temperature of the corona, and its emission is characterized by a half-million-degree temperature if we assume an ionization equilibrium as the source of the emission.

At times when the corona is more active, it shows a green line emission (Fe XIV, $E_p = 355$ ev) and the X-ray spectrum seems to be extended to shorter wavelengths and increased somewhat more intensely over all. At such times, without any evidence of solar flares, we have observed X-ray emission down to about 5 A.

It appears that this emission comes from the same regions that produce the green line. These are coronal condensations with temperatures of the order of 2 million degrees or perhaps higher.

The shortest-wavelength X rays shown in Figure 2 were observed during a small solar flare, class 1 minus. It was the first observation made from a rocket at the time of a solar flare. The X-ray spectrum extended down to a wavelength of about 3 A and could be characterized by a temperature of perhaps 4 million degrees.

At the base of the curve of Figure 2 are indicated the regions of the atmosphere in which the radiations are absorbed and in which they produce ionospheric electricity. The Lyman-α line comes down to the *D* region, about 75 to 90 km, under normal conditions, with the sun overhead. The broad range of X rays is effective in the *E* and *F* regions, and then as the X rays get harder they penetrate again into the *D* region and below.

This kind of information is suitable for the geophysicist, who is interested in how solar radiation interacts with the atmosphere and who does not always need fine detail in resolution of the spectrum. But for the astrophysicist, the ultimate goal, of course, is to get the complete line spectrum with high resolution.

Theoretically, this X-ray spectrum is composed of a continuum of which one-third is due to recombination radiation, and the remainder to line emission resulting from electron impact excitation, or emission following capture into excited states. Knowledge of the line spectrum would reveal the abundance of elements and the excitation conditions in the corona.

Figure 3 shows the type of spectrograph that has been used by Dr. Tousey and his associates at the Naval Research Laboratory for the past several years. It utilizes a normal incidence reflection diffraction grating. This particular instrument is designed to carry two spectrographs to cover the ranges 1500 to 3500 A and 500 to 2500 A, respectively. The focal length of each grating is 40 cm.

Figure 4 is a photograph of the pointing control developed by the University of Colorado to carry such a spectrograph. It is a biaxial control that permits the entire instrument to rotate

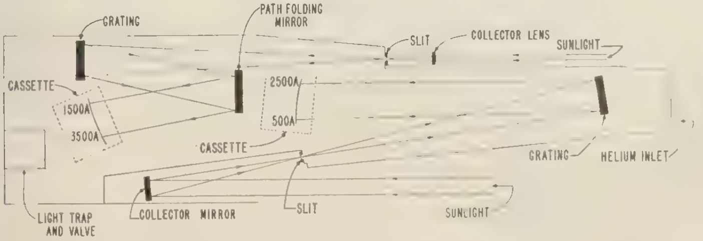


FIG. 3—Solar spectrograph used at the Naval Research Laboratory.

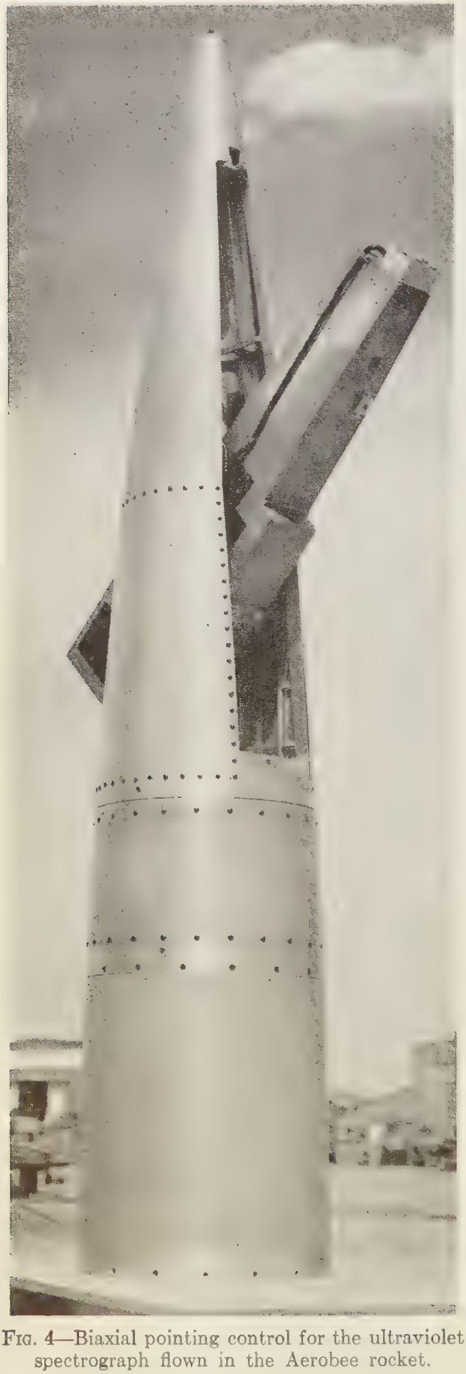


Fig. 4—Biaxial pointing control for the ultraviolet spectrograph flown in the Aerobee rocket.

about the long axis of the rocket, and the spectrograph itself is carried in trunnions so that it can swing about a transverse axis. In this way, in spite of a wide range of roll rates and yaw angles, the spectrograph is kept pointed at the sun throughout the flight. This is a very important piece of hardware for rocket astronomy, and only because the system works so well has it been possible to obtain the type of spectrogram shown in Figures 5 and 6.

Figure 5 is a spectrogram obtained very recently by the NRL group. Starting at longer wavelengths, the continuum has largely disappeared. The background is mostly stray light. The dark streak is the result of a dust particle in the slit system. The spectrum is characterized by emission lines originating in the chromosphere and in the corona, lines that belong to Si IV, C II, Si II, and other highly ionized atoms. The Lyman-α line shows up very brightly. Actually it is a narrow line, but the intensity is so great that halation spreads the image. Continuing down the spectrum, the Lyman-β line and the remaining lines of the Lyman series appear. Lyman γ is missing because apparently it is still very highly absorbed even at a height of 123 miles, from which this spectrum was obtained. A very weak emission line has been identified with MgX. It is interesting how far off theoretical predictions have been from the present observations. For example, the Mg X line was the strongest line in this part of the spectrum. The He I resonance line, at 584 A, is clearly present, though not very strong. A microphotometer trace of the spectrogram clearly shows the Lyman continuum.

This past year seems to have brought many major advances. Even earlier than the NRL experiment, Dr. Rense, at the University of Colorado, flew a grazing incidence spectrograph and was able to photograph the He II resonance line at 304 A. This line turns out to be very strong. His first impressions were that it was comparable in strength with the Lyman-α line.

More recently, Dr. Hinteregger, at the Air Force Cambridge Research Center, flew a novel type of spectrograph in which he used a Bendix photomultiplier tube that is almost totally insensitive to wavelengths above the Lyman-α wavelength but highly sensitive to shorter wavelengths. He used a spectrograph arrangement in

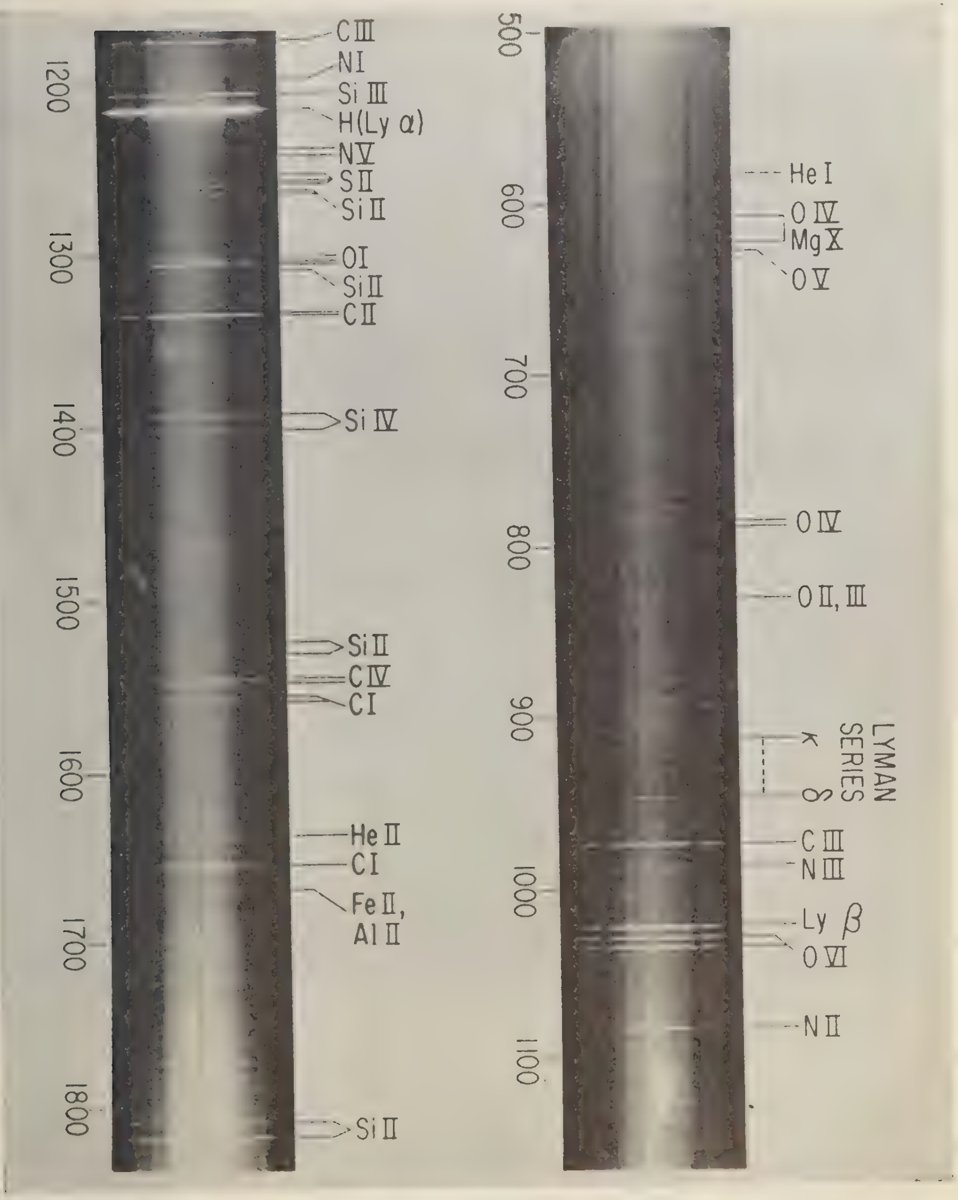


Fig. 5—Solar spectrogram obtained by NRL.

which the film was replaced by a scanning slit cut in a continuous belt. As this slit passed before the photocathode of the multiplier tube the counting rate was telemetered. This type of instrumentation will undoubtedly have great application in future work.

Dr. Hinteregger's first tentative estimate is that the helium resonance line at 304 A has an intensity of the order of a couple of tenths of an erg per square centimeter per second, which is very high compared with anything else in that part of the spectrum.

The most spectacular type of activity in the sun is the solar flare, and one of the primary purposes of the rocket astronomy program of the IGY was to try to get information on the X-ray and ultraviolet emission spectrum of a flare. Fortunately, about the time that this program was ready to start, important advances had been made in solid-propellant rocketry techniques. The two-stage solid-propellant system has been one of the most frequently used during the IGY, combining the Nike booster with a Deacon, a Cajun, or an Asp second-stage rocket. With this system the investigator can wait for a flare, which never gives any advance warning, and push the firing button on a moment's notice to get the rocket up within a minute or two of the beginning of a flare.

We succeeded in several attempts of this sort, and found that a flare produces a bright X-ray flash. In the largest flares the flash extended down to wavelengths of 1 to 2 A. The intensity of X-ray emission was great enough to produce the major effects associated with sudden ionospheric disturbances.

Once you have a spectrum, you would like to know where the radiations originate on the sun. Optical observatories carry on a continuous survey of the emissions from the solar surface with spectroheliographs at wavelengths of Hα and the calcium K line. We are all familiar with the types of photographs that result, showing evidences of turbulence and plage formations which statistically correlate with the behavior of the earth's ionosphere.

It is important to get pictures of the sun in Lyman α and the helium resonance lines and other ultraviolet emission lines originating in the chromosphere and corona. This of course is a very difficult thing to do, because conventional camera optics cannot be used in this part of the spectrum.

The first indication of success along these lines was obtained by Dr. Rense at the University of Colorado a couple of years ago when he tried to take a picture of the sun in Lyman α with lithium fluoride optics. Just a few weeks ago Purcell, Packer, and Tousey of NRL obtained a finely detailed picture by means of a mirror grating.

Figure 6 illustrates the camera used in the NRL experiments. It consists of two concave mirror gratings. The first one predisperses the

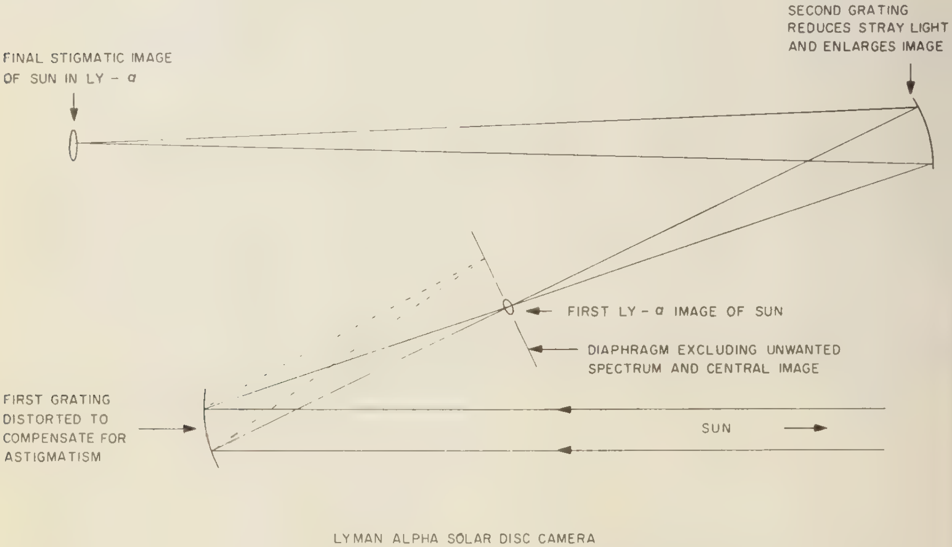
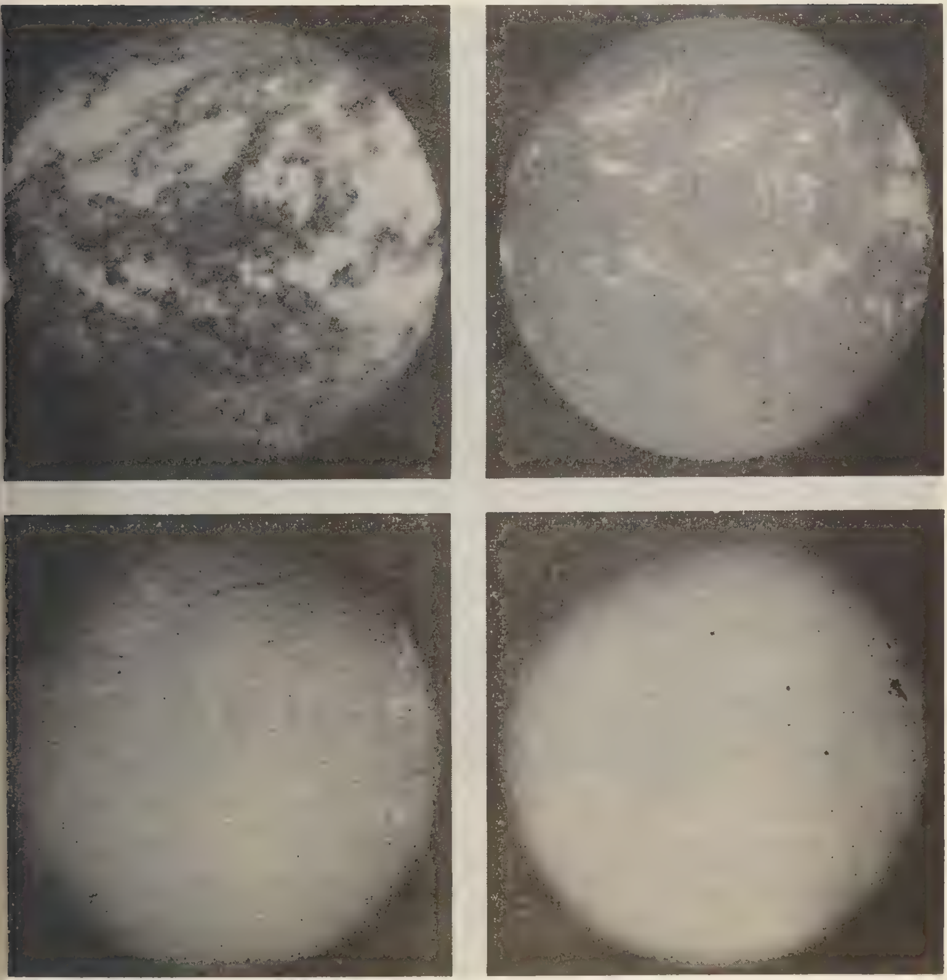

SECOND GRATING REDUCES STRAY LIGHT AND ENLARGES IMAGE
FINAL STIGMATIC IMAGE OF SUN IN LY - α
FIRST LY - α IMAGE OF SUN
DIAPHRAGM EXCLUDING UNWANTED SPECTRUM AND CENTRAL IMAGE
FIRST GRATING DISTORTED TO COMPENSATE FOR ASTIGMATISM
SUN
LYMAN ALPHA SOLAR DISC CAMERA

FIG. 6—Camera used to obtain pictures of the sun in the Lyman-α line.



IG. 7—Proceeding clockwise from the upper left, the four photographs of the surface of the sun are in Lyman α, calcium K, white light, and Hα.

nage of the sun in Lyman-α and other wavengths near by. A diaphragm with suitable perture is inserted to pass principally the yman-α image. Then a second dispersing sysm further refines the image and throws out the maining stray light. The final image has a solution of 20 seconds of arc.

The image of the sun in Lyman α (Fig. 7) as photographed in flight with resolution of at ast a minute of arc and possibly half a mine of arc.

There is a detailed correlation between the rocket picture of the sun in Lyman α, the calcium K plage picture obtained by the McMath-Hulbert Observatory and an NRL picture in Hα, all photographed at the same time. The Naval Observatory picture in white light shows the sunspot distribution. The bright Lyman-α plages occur over the sunspot regions just as the calcium plages do. The contrast is much greater in Lyman α than in Ca K and in the Hα. Over the course of a sunspot cycle there

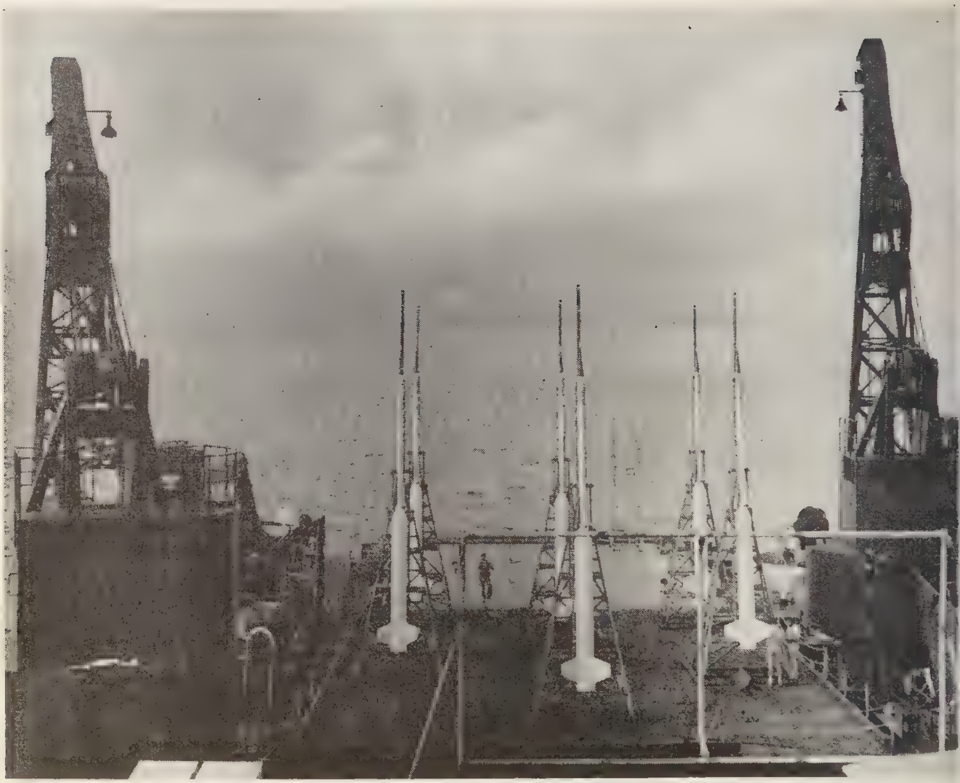


Fig. 8—Rockets used for measurements of the X-ray distribution from the sun during a solar eclipse

has been quite a wide range in Lyman-α intensity measurements with photon counters and ionization chambers. Judging by the appearance of the disk in Lyman α we should not be surprised by such variations.

If we try to extend this technique to other wavelengths, we must first achieve high reflectances on the mirror surfaces. Already considerable progress has been made in perfecting this Lyman-α camera. G. Hass of the Engineer Research and Development Laboratory at Fort Belvoir, in collaboration with Tousey's group at NRL, has developed coatings that yield 80 per cent reflectance at Lyman α. The exposure time was 1/50 second, and the transmittance of the camera for Lyman α was 11 per cent. It is possible now to develop coating surfaces for the ultraviolet which give promise of providing very high reflectivities down to quite short wavelengths. We can look forward eventuall to pictures of the sun in the helium resonanc lines. In the X-ray region of the spectrum th problem is much more difficult. X-rays canno be imaged by conventional mirror optics. Th simplest approach is a pinhole camera, but th intensities are far too low for available rocke exposures.

We tried to get a rough picture of the X-ra distribution from the sun by utilizing the various phases of the solar eclipse last fall in th South Pacific. Figure 8 shows the rocket arrangement used for the eclipse measurement Six Nike-Asp rockets were mounted on the helicopter deck of the *USS Point Defiance,* whic took us to the South Pacific along with the IG solar eclipse expedition. The plan was to fir these rockets at different phases of the eclipse s that we could see the occulting edge of the moo

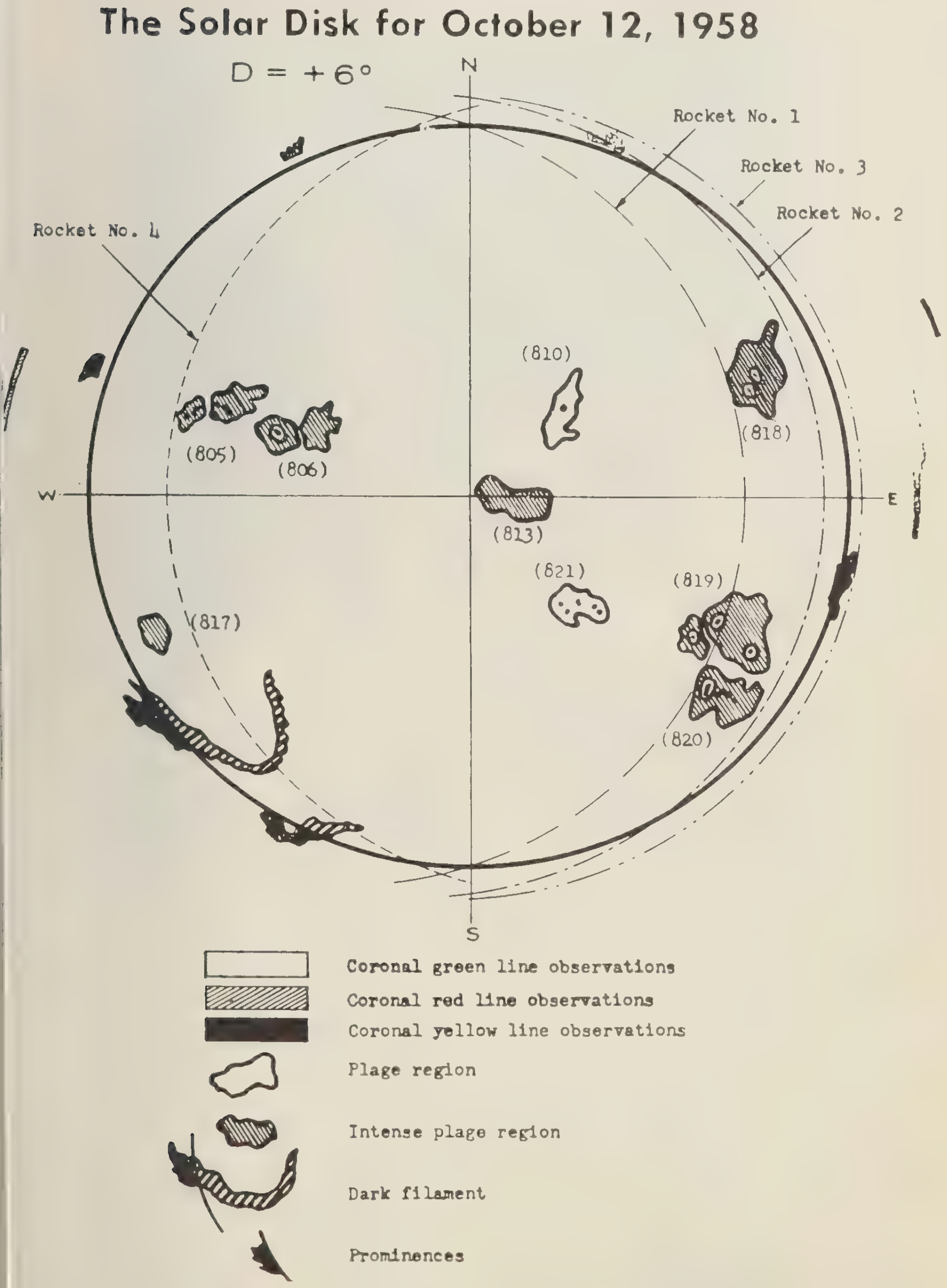


Fig. 9—Distribution of activity on the sun at the time of the solar eclipse.

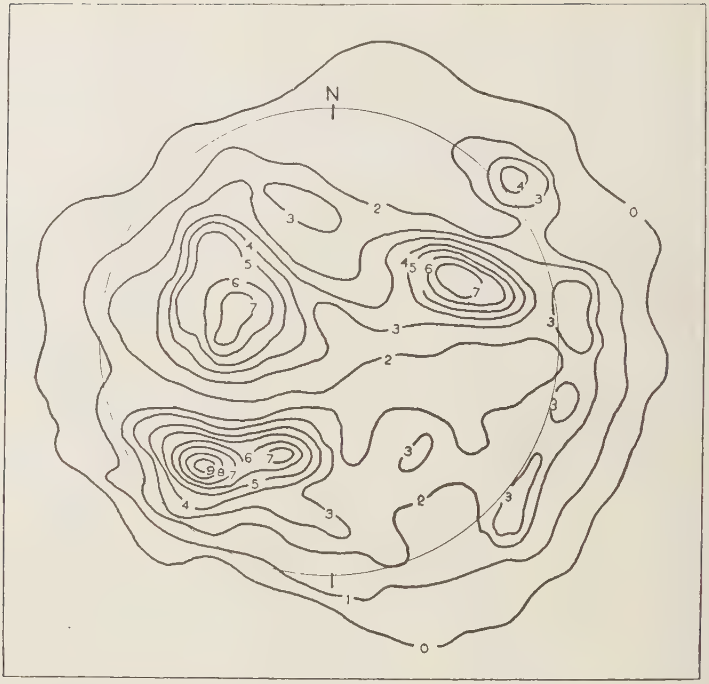


Fig. 10—Radio brightness distribution over the sun at 21 cm. The brightness temperature indicated o each isophote is in units of 10^6 °K (after W. N. Christiansen).

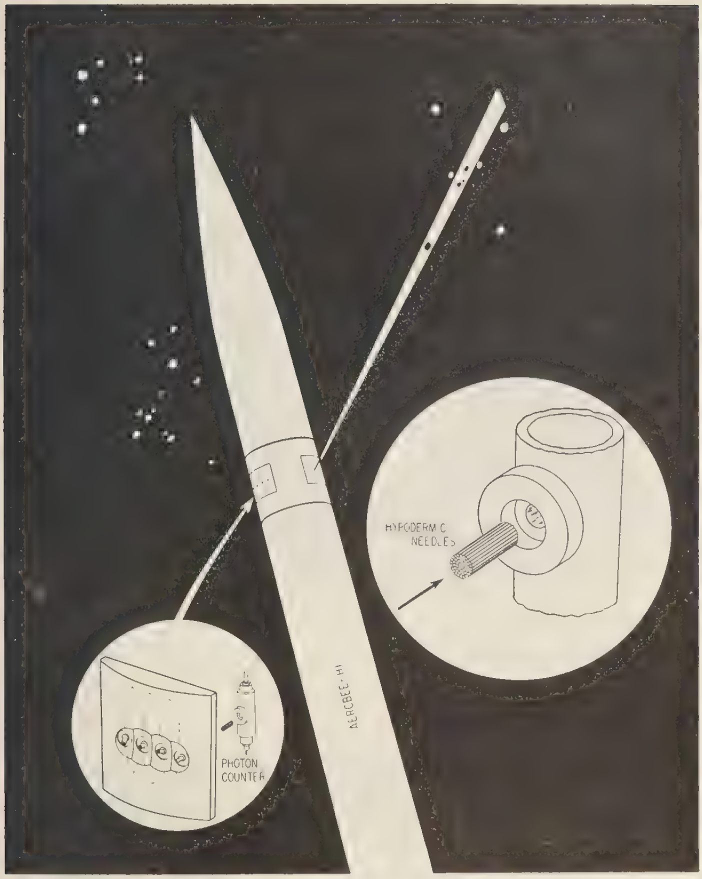
cross various active regions and also get measurements of the residual fluxes of ultraviolet, Lyman α, and X rays at totality. Figure 9 is a sketch of the distribution of activity on the sun at the time of the eclipse. The plan was to fire so that an east limb would be exposed on the first shot; then to fire two rockets during totality; finally to take an observation on the crescent of the west limb, which was relatively inactive compared with the east-limb crescent.

As a result of these measurements we found that the sun shows strong limb brightening in X rays, and that more than 25 per cent of the X-ray flux was residual at totality. In contrast, the Lyman-α intensity went down almost in direct proportion to the exposed area of the disk, leaving only 0.05 per cent residual at totality.

These results should have interesting consequences for theories of the ionosphere. It has been concluded from many measurements made with radio sounding techniques that there is a residual ionizing flux at totality. If this flux i strong, there must be a relatively higher recombination rate in the ionosphere. The rocke measurements indicate strong support for th X-ray theory of the source of *E*-layer ionizatio and a higher recombination rate.

Figure 10 is the sort of picture that has bee obtained of the sun in isophotes of radio brightness. I believe that if we were to photograp the sun in X rays the intensity isophotes woul follow a very similar distribution.

Rocket astronomy within the past two year has also given us the first picture of the ultraviolet night sky. Figure 11 illustrates the crud technique used thus far. In order to define th field of view, bundles of hypodermic needles ar placed in front of the photoelectric detector The collimation provided by these cylinders isolates a region about 3° in angular width in th sky. As the rocket spins, the collimators swee out a ring of the sky; and as the rocket yaw and precesses, and the plane of the ring tilts i



ıg. 11—Instrumentation used for the initial rocket astronomy experiments. Collimation is provided by the bundle of hypodermic needles.

ıe sky with a fortunate combination of roll and ıw the collimators can sweep a large part of ıe sky during the course of a flight.

Figure 12 shows an isophote map observed in ıe wavelength of 1300 ± 50 A in the direction ' Orion. Orion is a complicated region filled ith hot blue stars, dust, and visible gaseous ebulosity; also bright nebulosity is distributed ırough the Orion region in the ultraviolet.

Even though the detector bandwidth was only out 100 A, the surface brightness was of the der of 10^{-4} erg/cm^2/sec. It is surprising, first, ıat the ultraviolet nebulosity is so intense, and, cond, that it is much more extensive than the sible nebulosity, so that the total emission in ultraviolet is comparatively much greater than the visible nebulosity.

The situation around Spica in the constellation of Virgo is much simpler. When this star is viewed with an optical telescope, no visible nebulosity is apparent, but in the ultraviolet a very extended nebulosity, almost 20° in diameter, surrounds it (Fig. 13). The total emission from the ultraviolet nebulosity is comparable to the total black-body radiation below the Lyman limit, on the basis of a color temperature of 28,000°. This is a real puzzle, and the observations must be repeated with high resolution, both geometrically and spectroscopically, before we can begin to understand it.

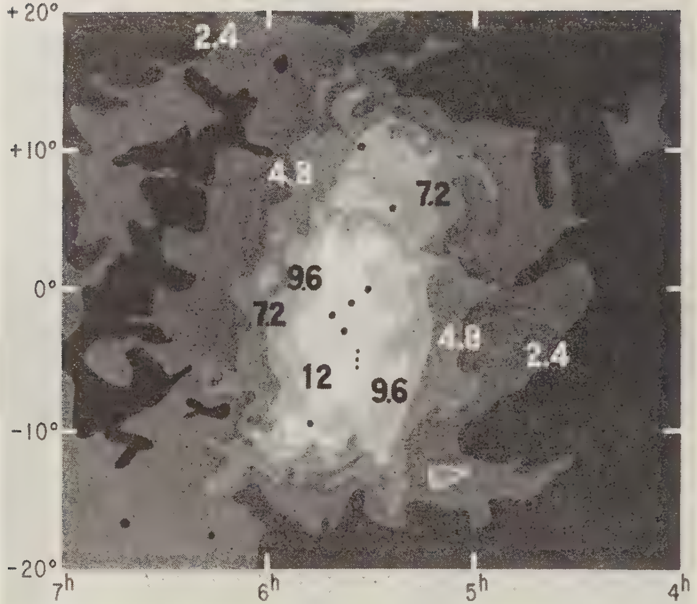


Fig. 12—Brightness distribution in the direction of Orion, at wavelength 1300 ± 50 A. The surfac brightness is indicated in units of 10^{-4} erg/cm²/sec. The dots mark the positions of familiar stars

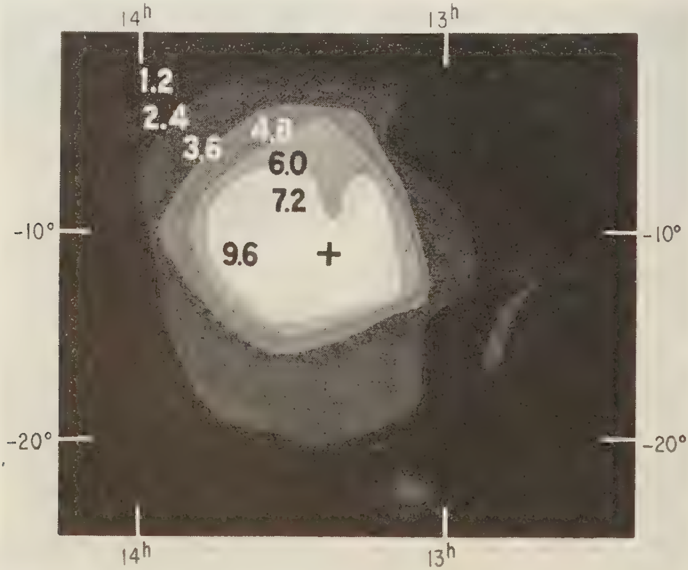


Fig. 13—Isophotes of the nebula around α Virginis. Surface brightness is indicated in units of 10 erg/cm²/sec. The cross marks the position of the star.

The ultraviolet nebulosities are superimposed n a bright diffuse glow of Lyman α that covers he entire sky. The most plausible explanation r the glow is that it is a resonance scattering f solar Lyman α which comes from the sun and scattered back to the dark side of the earth y neutral atomic hydrogen in space. The queson of the distribution of the hydrogen has not et been resolved. We first adopted the interretation that the hydrogen was interplanetary. To explain all the various aspects of the exerimental results, we had to conclude that the ydrogen was fairly cold. It is difficult, howver, to understand a hydrogen gas distributed the midst of solar winds, Biermann's particle treams, and the extended hot corona of the sun. ossibly the hydrogen we were looking at is ctually well beyond the earth's orbit, perhaps nother astronomical unit or two out in the ntisolar direction. Another explanation that has een offered is that the hydrogen is a geocorona. old has suggested that such a terrestrial corona xtending out to about 10 earth radii could acount for the observed results.

F. S. Johnson, at Lockheed, has actually reated the hydrogen as geocoronal and calculated its distribution. His model requires 10^4 neutral hydrogen atoms per cubic centimeter at the exosphere, falling off to 10/cc at 10 earth radii.

There are many interesting experiments that can be done to resolve this question further, and plans are under way to carry them out. One of the most important instruments to be developed is a high-resolution spectrograph to look at the Lyman-α line profile. Tousey, Purcell, and Packer, at NRL, have such an instrument about ready for flight. One of the problems that delayed its use was the fact that the speed was so low that it did not seem possible to photograph the Lyman-α line in the available exposure time. But to illustrate how important some of the auxiliary techniques are, the improvements in reflectivities of ultraviolet surfaces have speeded up the performance of this spectrograph by a factor of 60, which makes it possible now to achieve adequate exposure. The next step is to use photoelectric recording, which will permit studying the change in contour as a function of light. Such information can yield both the density and temperature distribution with height in a hydrogen geocorona.

Astronomy from Satellites and Space Vehicles

Leo Goldberg

University of Michigan Observatory
Ann Arbor, Michigan

Space experiments of interest to astronomy all naturally into three groups. First, the elimiation of the earth's atmosphere as a barrier o observation exposes to view the entire electromagnetic spectrum of radiation from extraerrestrial sources. It also permits the investigaon both of the faint outer extensions of the olar atmosphere, which are now obliterated by ne bright daylight sky, and of the weak radiaons from faint stars and nebulas which are masked by radiation from atoms and molecules ı the earth's upper atmosphere. Second, the dvent of satellites and space vehicles makes ossible a whole series of experiments and obervations which are absolutely unique and hich can test the foundations of physical theoes. In other words, space research introduces ne element of controlled experimentation into hat has always been fundamentally an obserational science. Finally, the investigation of the moon and the planets may now be carried out y direct exploration with space probes.

I shall mention only briefly a few examples f what we mean by controlled experiments in stronomy, and then pass on rather quickly to onsideration of the new knowledge that astronmers hope to gain by observation of the uniserse from outside the earth's atmosphere. The most exciting of the controlled experiments that ave been proposed so far are designed to test ertain predictions of the special and general heories of relativity. One prediction of the genral theory of relativity is that the frequencies f radiation which are emitted or absorbed by toms and molecules are slowed down in the resence of a gravitational field. This suggests hat an atomic clock in a satellite several hunred miles above the surface of the earth will ave a faster rate than its counterpart on the round. Another prediction is that a highly lliptical orbit of a body moving in a gravitational field will revolve slowly in its own plane at a rate that can be precisely calculated from the theory. The effect is known as the relativistic advance of the line of apsides. The line of apsides of a satellite orbit will rotate also for other reasons such as atmospheric drag and the oblateness and irregular shape of the earth, and the problem will be to disentangle these factors from the relativity effect.

Another experiment, proposed in several different forms by Dicke, Clemence, and others, would check the hypothesis that the time scales of gravitational and atomic clocks are slightly different. In a time interval short compared with the 'age of the universe,' the frequency of the atomic clock would increase with respect to that of the gravitational clock at a rate proportional to the time elapsed since the clocks were adjusted to the same frequency. In this experiment, the atomic clock would be one of the recently developed atomic or molecular masers, while, as proposed by Clemence, the gravitational clock would be a satellite of carefully selected design and orbit.

All the foregoing experiments, and many others, have become feasible not only because of the availability of satellite vehicles but also through the development of atomic time standards which are accurate to 1 part in 10^{10} or 10^{11}.

Let me now describe some of the astronomical observations that will be carried out from satellites within the next few years. The planning of such experiments two or three years in advance presents some rather serious difficulties, because the technology associated with space vehicles changes so rapidly that no one can be quite certain what the state of the art will be at the time a given experiment is scheduled. By the same token, the engineering problems that must be overcome before space observatories become a reality are so formidable that unless

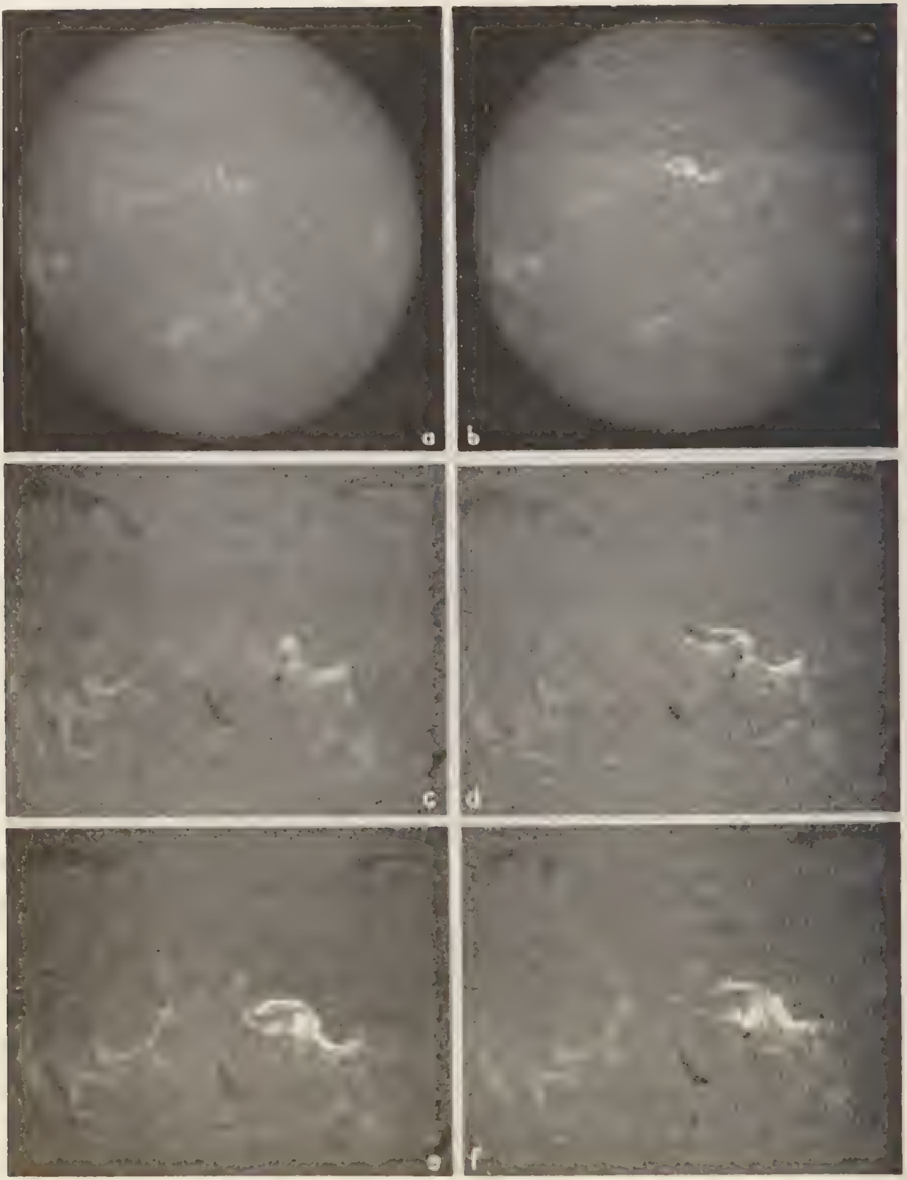


FIG. 1—Development of Hα flare, importance 3, 17°N12°W, August 22, 1958. In each picture north i at the top and east is at the left. *The McMath-Hulbert Observatory of the University of Michigan.*

(*a*) Entire disk, 12^h11^m UT, several hours before start of flare. (*b*) Entire disk, 14^h48^m UT, flare a maximum. (*c*) 14^h00^m UT, early stage of flare. (*d*) 14^h20^m UT, spreading of flare before principal ris to maximum. (*e*) 14^h48^m UT, flare at maximum. (*f*) 15^h21^m UT, post-maximum stage of flare. Note i (*e*) and (*f*) the extensive, low-intensity brightenings to the east (left) of the flare.

te astronomers make known their needs the sace vehicles may become available well in adnce of the instrumentation needed to equip em. The prospect now is that within about ree years an astronomical telescope can be put to an orbit with a radius of several hundred iles in a stabilized vehicle that can carry an strument payload of several thousand pounds. ven after allowance is made for the weight of é stabilization equipment, the power supply, e telemetering apparatus, etc., an order of agnitude of half a ton in weight of instruments ll be available. Many instrumental problems ust be solved before this prospect becomes a ality. The satellite telescope must not only be ablized so that it will point accurately in any sired direction in space, the direction of pointg must also be controllable by command from e ground. The desired stabilization will probbly be achieved by means of gas jets and rotatg flywheels. An adequate power supply must provided, with reasonably long lifetime. The chnique of using solar cells to recharge small storage batteries has been notably successful in Vanguard I for a full year of operation, although the power level is tiny. There seems to be no reason, however, why the same technique could not be employed on an expanded scale to produce 10 to 20 watts of power, which would be quite satisfactory at present. Radioactive sources of power and chemical fuel cells also offer promise for the future. Most of the problems connected with the storage of data and their transmission to the ground by telemetry have already been solved in connection with earlier rocket and satellite experiments. There still remains, however, the extremely difficult problem of designing optical systems and detectors that will operate with high efficiency and for long times for radiation of very short wavelength.

It is generally agreed that the sun will have very high priority as an object for investigation from space vehicles. It is the only star whose detailed surface features can be studied, and its radiation has a very direct and practical influence upon our daily lives. Every advance

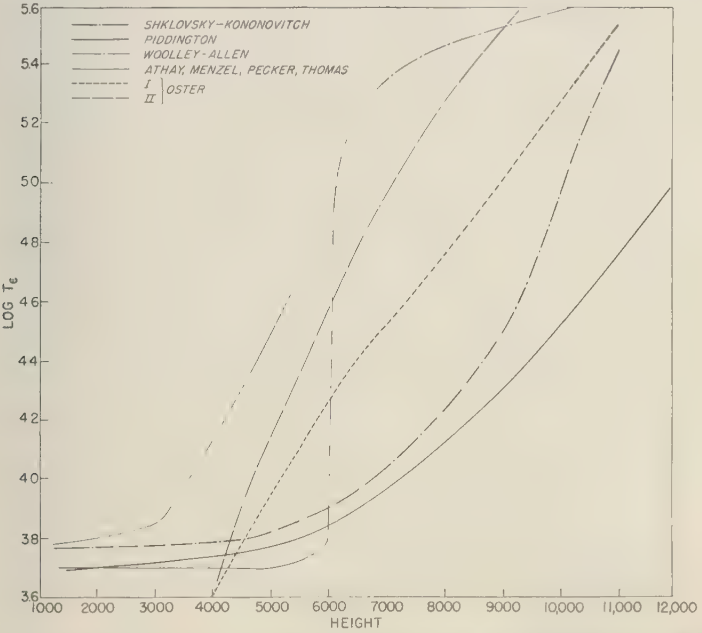


FIG. 2—Temperature distribution in the solar chromosphere as derived by different investigators. The garithm of the electron temperature is plotted against the height in kilometers above the photohere.

made toward the understanding of the sun serves also to expand our knowledge of the other stars. Although a vast amount of knowledge has been accumulated about the sun, many of the most fundamental questions concerning its structure and activity are completely unanswered. Our knowledge is most deficient with respect to the outermost layers, the chromosphere and corona, both as regards their structure and the nature and origin of the often catastrophic disturbances that take place within them. The most energetic of these disturbances is, of course, the solar flare (see Fig. 1), in which huge quantities of energy, sometimes exceeding 10^{30} ergs, are released in a relatively small region of the solar atmosphere during the space of a few minutes. Figure 2 illustrates the very great range of disagreement among different investigators who have attempted to derive the temperature variation with height in the solar chomosphere [*Aller and others*, 1958].

One reason for our poor understanding of the sun's outer layers is the very complexity of their structure and of the events that take place within them. A more fundamental reason is that most of the radiation from these layers occurs at the two extreme ends of the electromagnetic spectrum. For example, observations from rockets and balloons have revealed that ultraviolet radiation from the sun is enhanced during solar flares, and rather intense bursts of X rays and γ radiation have also been observed. It now appears probable that the X-ray bursts in particular are responsible for the sudden disturbances in the earth's ionosphere which result in short-wave radio fadeouts [*Friedman and others*, 1958]. Bright flares are almost always accompanied by great bursts of radio noise, particularly in the low-frequency band from 20 to 600 Mc/sec (see Fig. 3).

Many interesting experiments are being planned for the observation of X rays and ultraviolet radiation from the sun. Since the intensity of the radiation is time dependent, the experiments should be done from satellites as well as from rockets. Perhaps the most urgent experiment is the monitoring of the spectrum with scanning spectrometers. This experiment need not await development of refined stabilization techniques since a great deal can be learned from analysis of radiation averaged over the whole disk of the sun. As soon as accurate pointing capability is acquired, small regions on the surface of the sun can be isolated for more detailed studies.

The photographic technique has been used with extraordinary success by Tousey and his associates at NRL and by Rense at the Uni-

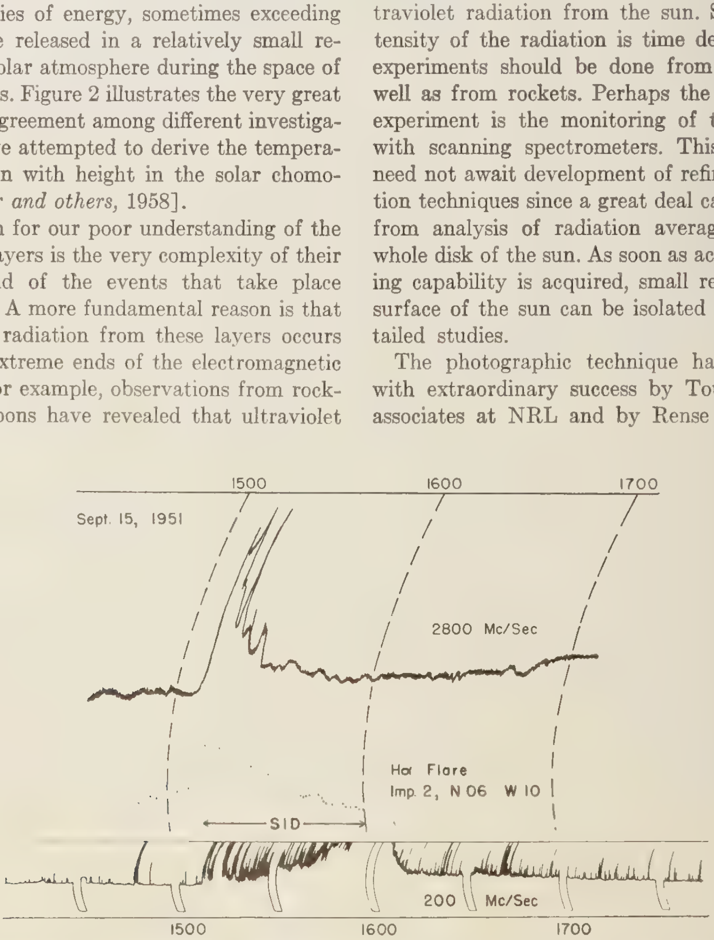
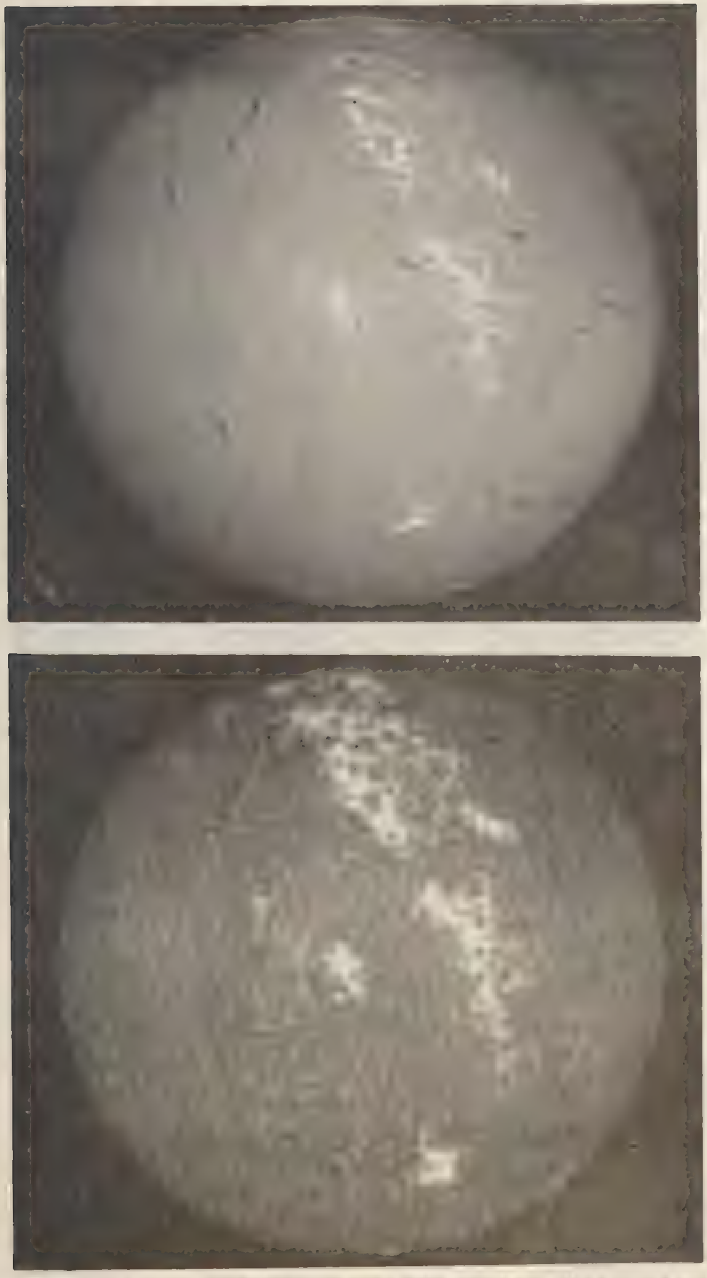


FIG. 3—Comparison of photometric light curve of an Hα flare, September 15, 1951, with concomitant radio-frequency emission. The 2800 Mc/sec record was obtained by Covington at the National Research Council, Canada, and the 200 Mc/sec record at Cornell University.



4—Hα and K_2 spectroheliograms of the sun, May 1, 1958. North is at the top and east is to the left of each photograph. *The McMath-Hulbert Observatory of the University of Michigan.*

versity of Colorado in recording the solar spectrum from rockets. For satellite use, however, the technique suffers from two disadvantages: there is no way at present of recovering a film from a satellite; and photographic emulsions are vulnerable to light scattered from the visible region of the solar spectrum. Both disadvantages appear to have been overcome by J. Hinteregger (private communication, 1959), of the Cambridge Air Force Research Center, who has recently flown a 2-meter grazing incidence spectrograph with a so-called magnetic photomultiplier as energy detector. The detector appears to be completely insensitive to radiation of wavelength longer than about 1200 A. During the flight of the rocket to a height of 210 km, the spectrum from 200 A to 1000 A was successfully scanned and the intensities telemetered back to ground. This technique would appear to be very well suited to satellite application.

A second series of experiments would be concerned with recording images of the entire sun in monochromatic radiation at ultraviolet and X-ray wavelengths. Similar monochromatic photographs made in visible light from the ground (see Fig. 4) have already constituted a major source of information on the low chromosphere, particularly with respect to the transient disturbances. An observation of very great significance would be the simultaneous recording, during a solar flare, of solar images in the light of the Lyman-α line of neutral hydrogen at λ1216, the λ584 line of He I, and the λ304 line of He II—all resonance lines. Initially, a great deal could be learned with an angular resolution of about 1 minute of arc. With more sophisticated stabilization and an angular resolution of 1 or 2 seconds of arc, such monochromatic images would greatly advance our knowledge of the structure of the chromosphere, especially with reference to its inhomogeneities. Very successful spectroheliograms of the sun in Lyman-α radiation have also recently been obtained by Tousey from a rocket.

An interesting by-product of the solar-spectrum studies would be the determination of the densities of the constituent atoms at altitudes above 200 km in the earth's atmosphere, as suggested by *Spitzer* [1956]. The upper atmosphere of the earth contains neutral atoms of hydrogen, oxygen, and nitrogen. Emission lines of these elements are emitted by the solar chromosphere and, during times of sunrise and sunset as seei from a satellite at point *S*, Figure 5, they woul be absorbed by the earth's atmosphere. Measurements of the intensity variations in Lyman- λ1216 of hydrogen, in the triplet of oxygen nea λ1300, and in the triplet of nitrogen near λ113 could lead to precise determinations of the density gradients of these constituents of the earth' atmosphere.

Turning now to the other extreme end of th electromagnetic spectrum, we are interested i the observation of radio waves at frequencie less than about 20 Mc/sec, which are reflecte back into space by the earth's ionosphere. Special interest attaches to the observation of sola radio bursts in the very low frequencies. The dynamic spectra of these bursts (see Fig. 6), whic originate high up in the chromosphere and i the corona, have been observed to a low-frequency limit of 40 Mc/sec by *Wild and othe* [1954]. The so-called type II and type II bursts, which begin at high frequencies and occur progressively later at the lower frequencie have been interpreted by Wild as resulting fro corpuscular streams propagated outward throug the solar atmosphere at speeds varying from few hundred to a few tens of thousands of kilometers per second.

It is suspected that some of the corpuscul streams have sufficient kinetic energy to escap from the solar atmosphere and to reach th earth, whereas others, of lower energy, are turne back or stopped. To establish whether the rad bursts are caused by the same corpuscul streams that are also responsible for geomagnet storms or for the very soft component of sol cosmic rays, F. T. Haddock has proposed th the low-frequency end of the solar burst spect be observed from a satellite down to 1 Mc/s or less. He is at present engaged in designi such an experiment, which may also yield as by-product the decrease of electron density wi distance from the sun and perhaps the accelertion or deceleration of the corpuscular stream the initial phase of its flight from the sun to t earth.

The stars and nebulas—Satellite investigatio of the stars, the nebulas, the interstellar mediu and external galaxies are at least an order magnitude more difficult experimentally th

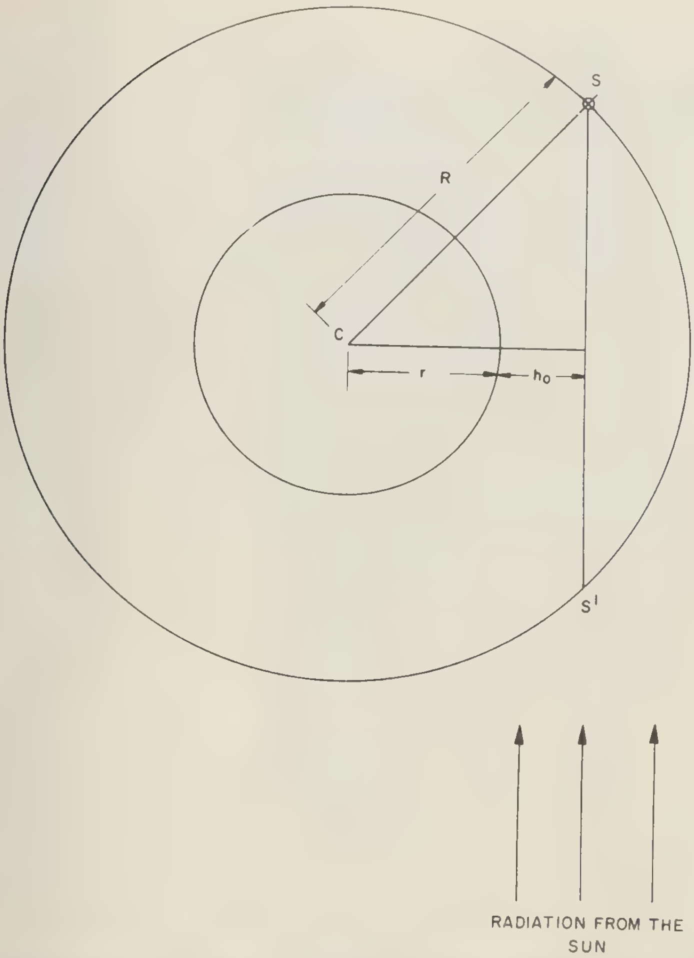


Fig. 5—Atmospheric absorption experiment. Radiation from the sun passes through and is absorbed
the atmosphere along the direction SS' before reaching the satellite at S. The absorption is deter-
ined chiefly by the density at height h_0.

.ose dealing with the sun, owing to the relative
intness of the sources and the stringent re-
uirements for accurate guidance and control.
urthermore, astrophysical measurements from
ckets have to date been limited almost en-
ely to solar spectroscopy, and therefore the
lactic experiments are not as well defined as
e solar experiments. For these reasons, galactic
d extragalactic research from satellites will
obably proceed in two stages. It is extremely
portant as a first step that the ultraviolet
radiation from the night sky be mapped quickly, even with relatively low angular and spectral resolution. A start in this direction has already been made from rockets by Friedman and his group at the Naval Research Laboratory, with some rather unexpected results. It has been found, for example, that the entire night sky is aglow with diffuse Lyman-α emission which appears to be solar radiation backscattered by cold neutral hydrogen either in the interplanetary medium or in a cloud surrounding the earth.

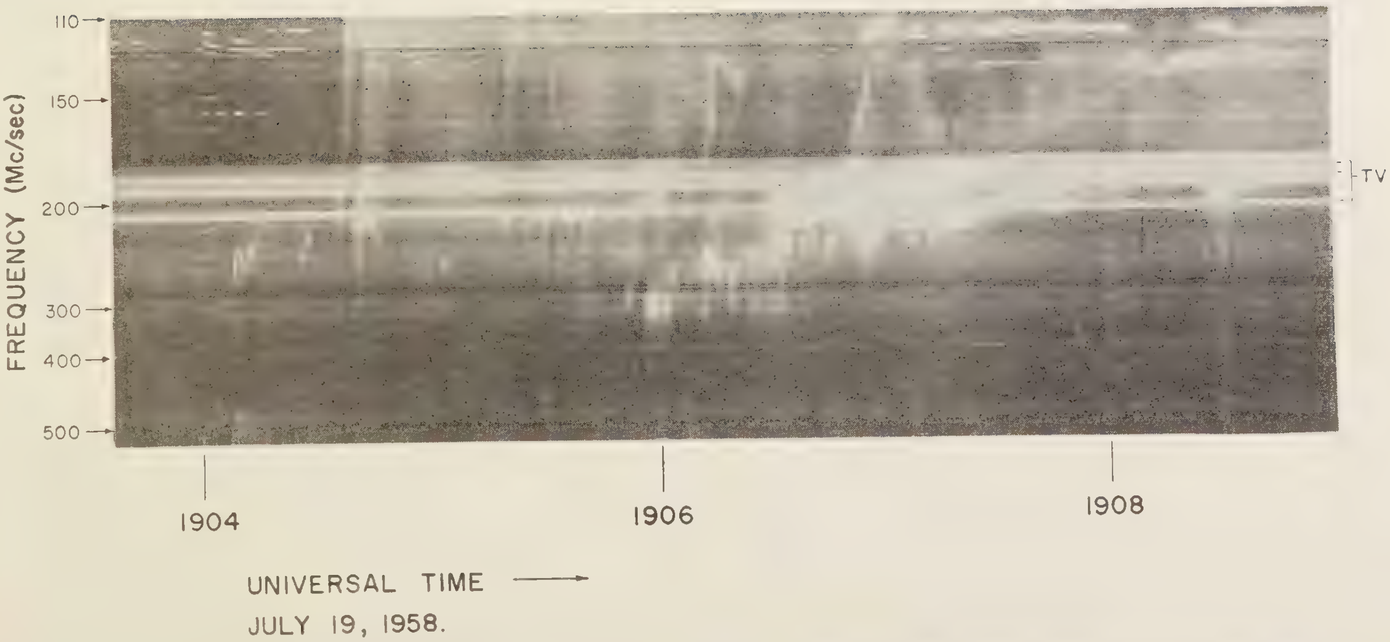


FIG. 6—A low-intensity modulated display, recorded at the University of Michigan, of the dynamic spectrum of an unusual solar radio burst complex at the time of a solar flare. The horizontal bright lines, both solid and broken, are man-made interfering signals. The vertical columns of dashes are the frequency and time markers. The bright vertical lines, fuzzy brightenings, and curved increases in brightness are due to radio emission from the sun.

urthermore, radiation in a wavelength band entered at about 1300 A has been detected vith very high intensity from several extended ources, some of which coincide with visible ebulas and others which do not. The physical nechanism that produces this radiation is not nown.

A number of astronomers, including W. A. 3aum, A. D. Code, L. H. Aller, and F. L. Vhipple, have proposed experiments for comrehensive sky surveys designed to yield measrements of the radiation intensity at a number f wavelengths and with varying degrees of ngular and spectroscopic resolution. As a beinning, the observations could well be made ith a band pass of about 100 A width and ith an angular resolution between 1° and 5°. very effort should be made to obtain relative ntensities, not merely from one part of the sky another, but from one wavelength to another. f the wavelengths are properly chosen, valuable nformation could be obtained on the physical ate of interstellar matter, on the ultraviolet nergy distribution of bright, hot stars, and on ne relative contributions to the integrated radiaon background of starlight, galactic light, exagalactic nebulas and zodiacal light.

The wavelengths of the survey should inude a band below the limit of the Lyman eries in order to evaluate the importance of nterstellar absorption as a factor limiting the bservation of ultraviolet galactic radiation. everal astronomers have raised questions as to hether the interstellar absorption of neutral ydrogen is not so great as to screen effectively ot only Lyman-α radiation from even the ear-by stars but also all radiation for several undred angstroms below the Lyman-series mit at 912 A. These doubts have been strengthned by some recent illustrative calculations *Aller* [1959]. Consider first the Lyman-α ne, and the probability of its detection as an mission line in the gaseous nebulas or in the ight line stars. More specifically, consider the gion of λ Orionis in the Orion complex for hich *Wade* [1958] has found a neutral hydron abundance of about 10^{21} atoms per square ntimeter in the line of sight. If a temperature 100°K is assumed for the neutral hydrogen s, the optical thickness at the center of the ne is about 10^9 and radiation damping alone causes the line to be totally black over a width of about 7 A. A line-of-sight velocity of about 700 km/sec would be required to displace an emission line before it might become visible. Further, if the average concentration of neutral hydrogen atoms in the neighborhood of the sun is taken as $0.1/\text{cm}^3$, even the optical depth to Sirius in the center of Lyman α is about a million, and the half-width of the interstellar absorption is about 0.4 A.

Absorption in the Lyman continuum of the interstellar gas and also by neutral helium may present an even more serious problem. Again using the region of Orion as an example, the optical depth would be 6000 at the Lyman limit. The optical depth at a wavelength of 20 A is still about 0.4, and at 10 A about 0.1. It may thus be that observations of ultraviolet stellar radiation will, except for some of the nearest stars, be confined to X rays and to wavelengths longer than 912 A.

After the sky surveys, the next stage of galactic investigation would come with the establishment of accurately controlled telescopes in satellites. Two groups are already at work on the design and development of such telescopes, at Princeton under Lyman Spitzer and at the Smithsonian Astrophysical Observatory under F. L. Whipple. The Smithsonian group plans to begin with a telescope of about 8-inch diameter together with associated television equipment that will provide an image of an area of the sky in a monitoring station on the ground. This would be used for an ultraviolet sky survey, and an objective prism and grating would also give low-dispersion spectra. Spitzer envisages a somewhat larger telescope, perhaps up to 24 inches in diameter, to be used in conjunction with a high-dispersion stellar spectrograph for investigation of individual stars and the interstellar medium. A longer-range program looking toward the development of a relatively large satellite telescope has been proposed by A. B. Meinel and the the staff of the Kitt Peak National Observatory at Tucson.

Of special interest for investigation by ultraviolet stellar spectroscopy would be the chemical abundances of the elements in stars and in the interstellar medium; their precise determination is crucial from the standpoint of stellar evolution. A great deal has been learned about

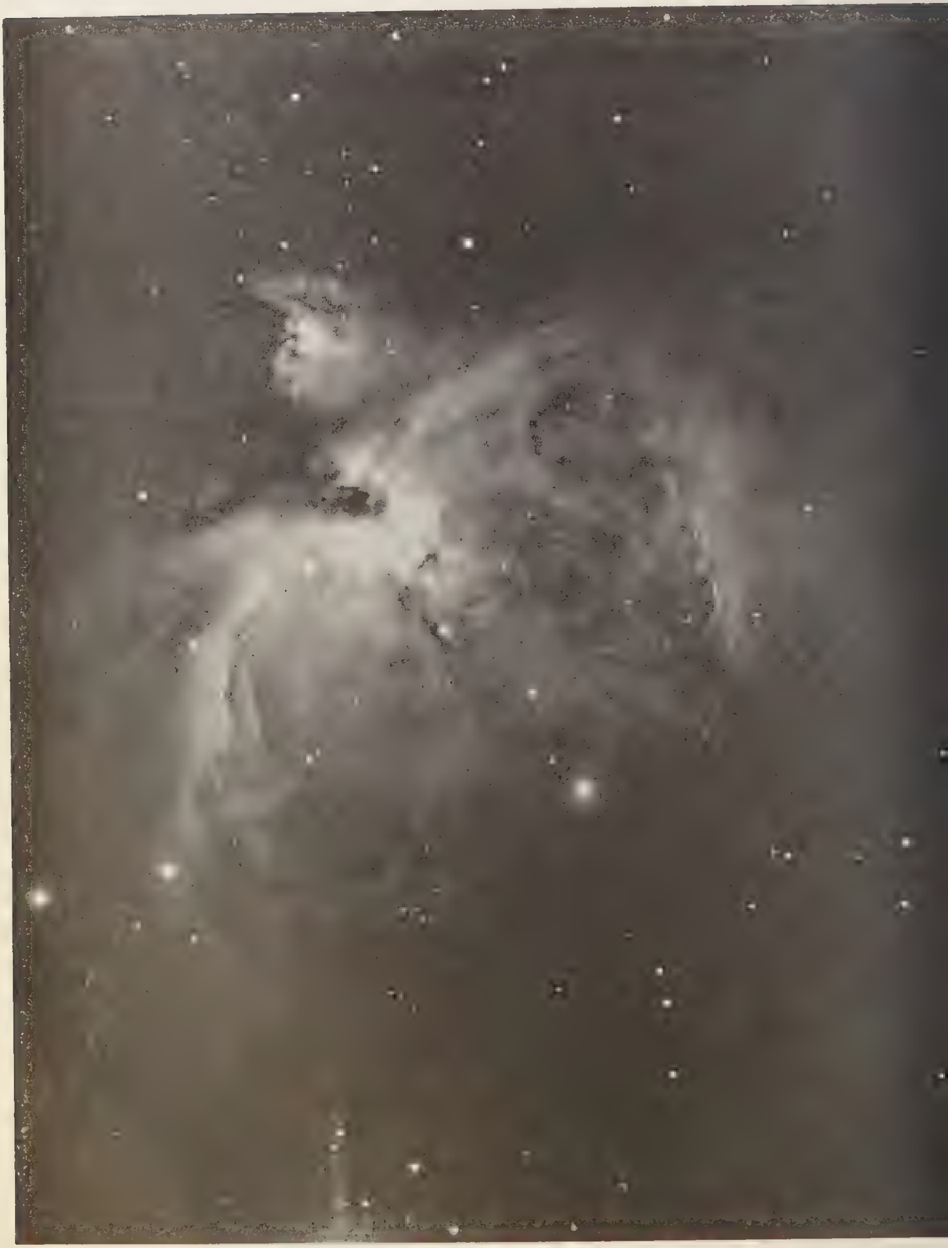


Fig. 7—The great gaseous nebula in Orion. *Mount Wilson and Palomar Observatories.*

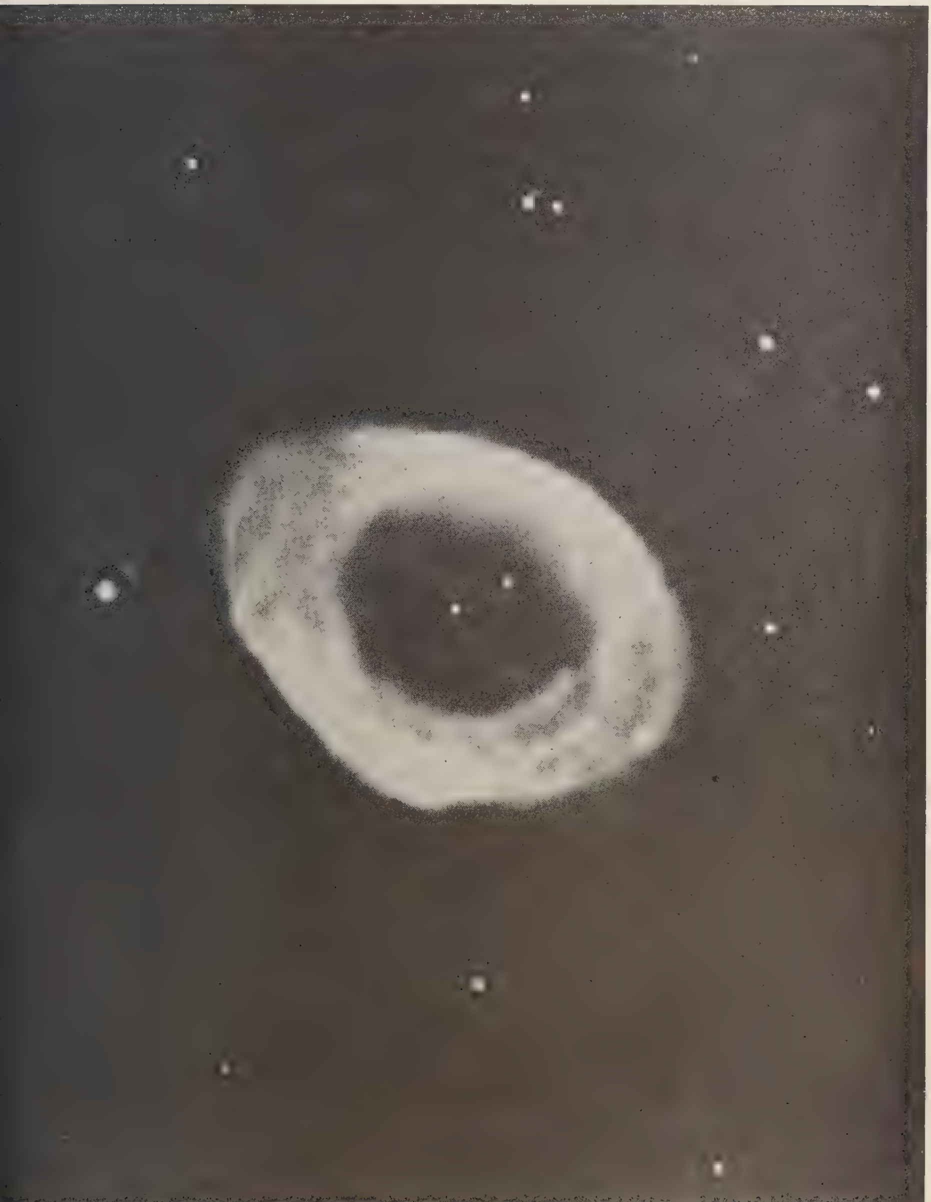


Fig. 8—The Ring planetary nebula in Lyra. *Mount Wilson and Palomar Observatories.*

ement abundances from observations of the isible spectrum. However, most elements are epresented in the visible spectrum by lines riginating from excited energy levels, and for nem the abundance determination rests upon ne application of a very large and uncertain emperature correction. The fundamental, or esonance, lines of such critical elements as hyrogen, helium, carbon, nitrogen, oxygen, and nagnesium fall in the ultraviolet region of the pectrum. Furthermore, some elements among ne inert gases and the halogens, such as neon, rgon, and fluorine, although well represented ι the earth, meteorites, nebulas, and certain eculiar stars, have no detectable lines in the isible region of normal stellar spectra. Thus, ur knowledge of element abundances in stars nd interstellar gas would be greatly improved y the advent of satellite spectroscopy. As was ointed out earlier, however, the abundance udies may be thwarted, at least in part, by ne absorption of the interstellar medium.

A highly fruitful area for investigation from abilized satellite vehicles would be the many ypes of emission nebulas that are found in the alaxy. Some, like the great nebula in Orion Fig. 7), are irregularly shaped objects associted with stars of very high surface temperature nd are indeed excited to fluorescence by the ltraviolet radiation emanating from these eighboring stars. The symmetry of the planetary ebulas (Fig. 8) suggests that they are highly xtended outer atmospheres of the central stars hose ultraviolet radiation again excites the as by fluorescence. Another type of gaseous ebula seems to derive its excitation from vionet internal turbulent motions, the collisions etween turbulent elements generating shock aves. The best-known example, the Crab nebla (Fig. 9), seems to be the after-effect of an ncient supernova explosion. The systematic ivestigation of the ultraviolet emission line pectra of nebulosities of these types would put n a firm basis our notions about the radiation alance, electron density, electron temperature, c. Highly turbulent gases like those in the rab nebula may emit fairly strongly in the -ray region, and, in fact, it is not unlikely nat a systematic survey of the sky for sources f X radiation would reveal an entirely new ass of objects as occurred in our recent experience in radio astronomy. It is particularly important that the extended regions in the sky discovered by Friedman and his collaborators to emit extremely large amounts of ultraviolet radiation in the neighborhood of λ1300 be subjected to detailed spectroscopic analysis.

The most important immediate application of the techniques of radio astronomy to the investigation of galactic and extragalactic sources will undoubtedly come at frequencies below 20 Mc/sec, which are now turned back by the ionosphere. Several groups have proposed to make observations of this kind from satellites for the following reason. The background of cosmic radio noise observed on the ground at low frequencies is compounded of contributions from galactic and extragalactic sources and from both thermal and non-thermal processes. Identification of the separate components of the cosmic background radiation is important from the point of view of cosmology, galactic structure, and other aspects of astrophysics. It is probable that the thermal and non-thermal emission processes that contribute to the background spectrum produce a maximum brightness temperature at some frequency in the band 1 to 20 Mc/sec, with a rapid decrease at lower frequencies. The observation of the low-frequency end of the spectrum from a satellite would be designed to reveal the percentages of the thermal and non-thermal contributions from both galactic and extragalactic sources.

The first few years of satellite astronomy will probably be devoted to the astrophysical analysis of radiation. As the technology of instrumentation advances, however, it may be expected that there will be important applications in the field of astrometry, which is concerned with the precise measurement of the positions and motions of the stars. P. van de Kamp and H. Eichhorn have communicated a few preliminary ideas on the advantages that would accrue to astrometry if a telescope could be mounted above the atmosphere. Three important sources of systematic errors would be obviously eliminated, namely, imperfect transparency and seeing and atmospheric dispersion. Images would be smaller, and fainter stars would be reached, an additional great advantage being the attainment of high optical resolving power in the ultraviolet. Conditions would be particularly favorable for observing faint white dwarf stars, whether single stars or components of binary stars. A survey for binary stars among 'single' stars would be more penetrating than from the bottom of our atmosphere, and in particular it should be easier to discover faint companion stars and possible planetary companions. The possibility of discovering many faint red dwarf stars would also be enhanced in the absence of night-sky emission in the red and infrared regions of the spectrum.

References

Aller, L. H., *Pubs. Astron. Soc. Pacific,* in press, 1959.

Aller, L. H., L. Goldberg, F. T. Haddock, and W. Liller, Astronomical experiments proposed for earth satellites, *Univ. Mich. Research Inst., Final Rept.* 2783–1–F, McDonnell Aircraft Corp., 1958.

Friedman, H., T. Chubb, J. E. Kupperian, R. W. Kreplin, and J. C. Lindsay, *IGY Rocket Rept. Series 1,* 183, 1958.

Spitzer, L., Jr., Earth satellites as research veh cles, *Franklin Inst. Pa. Monograph 2,* p. 69, 195

Wade, C., thesis, Harvard, 1958; quoted by T. Menon, *Proc. IRE, 46,* 232, 1958.

Wild, J. P., J. D. Murray, and W. C. Row *Australian J. Phys., 7,* 439, p. 10, 1954.

Discussion

Question: Is it certain that the interstellar m dium is neutral hydrogen, or is it possible that may consist of protons?

Mr. Goldberg: There are some regions of t interstellar medium that are ionized, but we n have a direct method of determining the quanti of neutral hydrogen along the line of sight various directions by measurement of the 21-c radiation. So these calculations were based on actual measurement.

Mr. Spitzer: It has been asked whether hydrog could be neutral or ionized. It could also be mole ular. I think the biggest gap at the present tin in adding up the mass budget of the universe the question of the amount of molecular hydr gen. To my mind one of the chief interests of spectroscopic satellite is that we can hope to dete the presence of molecular hydrogen directly.

Experimental Research Program in the Space Sciences

J. W. TOWNSEND, JR.

National Aeronautics and Space Administration
Washington, D. C.

The present paper is offered as an illustrated ıvelogue through new areas in satellite and ace probe research that are now under conleration by the National Aeronautics and ace Administration.

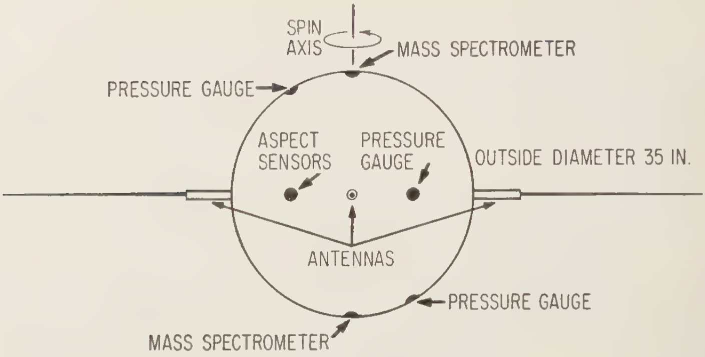
First of all, the devices for this research are ond-generation objects, that is, payloads of struments that are approximately 10 times ger than those flown in the Vanguard, Exorer, and Pioneer programs, but still smaller an the very large payloads being planned r such projects as the orbiting astronomical servatory. I should like to emphasize that e instrumentation for these various devices is ing created at many places, in universities, private industry, and in government laboraries. I shall not always indicate who has reonsibility for the instrumentation, because t all the assignments are firm.

An earlier paper by Dr. Newell outlined e program of the NASA in the space sciences. is program divides into certain disciplines or eas, namely: atmospheres, energetic particles, ospheres, astronomy, fields, and biology. In e present review I shall state the criteria for ace instrumentation design and discuss one ample of a NASA-sponsored scientific payload each of these fields of research except logy.

First, I should like to remind you that the al problem is one of systems. The 'systems' ncept must go over into the science program ause these devices are very complicated m a technological standpoint. The compliions are illustrated by a chronological acunt. In the first phase we must establish a ogram of research, then decide on certain jects to pursue, and finally conceive devices satisfy the objectives of the program and various projects. The second phase constitutes research and development on the various parts of the device, such as the instrument that is making the actual measurement, the sensor, or the end organ. Telemetering equipment must be considered, as well as other electronic devices such as command receivers and timers, and the various specialized gadgets that make it possible to program the measurements. We must also consider the power supply, an important part of the system, particularly if we wish to make measurements over long periods of time. We then enter a phase in which we worry about the mechanical structure of the object and the tests to be performed on it. The testing must be thorough, because the instrumentation cannot be adjusted or repaired in space. Next, we must integrate the device into a vehicle system that will provide the energy for placing it in orbit or freeing it from the earth's gravitational field. It is necessary to track the device and determine its position in space. We must telemeter and receive the data, remove them from the magnetic tapes, and convert them to a form suitable for analysis and final interpretation.

A few remarks may be made about the instrumentation itself. In order of importance the criteria for the various scientific devices are as follows: (1) reliability, the attribute of greatest importance; (2) ruggedness; (3) weight; (4) size; (5) sensitivity; (6) accuracy; (7) reproducibility; and finally (8) cost. The high cost of space instrumentation is always a limiting consideration in space research. For example, the total cost for the unmounted components of the solar cells for the power supply for one space payload may exceed a million dollars.

A study of these factors has already been completed in each of the areas of research that



NASA

Fig. 1—Schematic representation of the atmospheric structure satellite.

I mentioned, and this effort has led to five new payload designs which will be reviewed below. These designs are only preliminary, although for one or two the work has progressed considerably. The size of the instruments is not indicated in the accompanying figures, but these devices have dimensions of the order of 1 meter, intermediate between those in Project Vanguard and giant devices such as the orbiting observatories.

Atmospheric structure satellite—The first device I am going to talk about is an atmospheric structure satellite, designed for the measurement of the distributions of pressure, temperature, and density in the earth's high atmosphere, in particular these distributions as functions of geographical position, altitude, and time.

In this satellite we plan to fly mass spectrometers and pressure gauges as shown in Table 1. The mass spectrometers will be of a new design now under development, in which mass separation is effected by a magnetic field in place of the radiofrequency field used in the rocket program up to the present. The pressure gages are cold-cathode Redhead gages. They resemble the Phillips gage and are sometimes referred to as "magnetron" gages. We also have two hot-cathode ionization gages, of the Baird-Alpert design. The aspect sensors will determine the attitude of the satellite at the particular instant that we are making the measurement i.e., the orientation of the gage relative to the wind in the atmosphere at the time of the measurement. The telemetering and tracking transmitters are included in features of practically all these devices. Some of these instruments include hot filaments with a heavy power consumption; hence we can only make th measurements for brief periods and therefor require the command receiver. The weight o this satellite is approximately 300 pounds.

Table 1—*Instruments for the atmosphere structure satellite*

Two mass spectrometers
Two cold-cathode total pressure gauges
Two hot-cathode total pressure gauges with ion traps
Aspect sensors: magnetic and optical types
Telemetering transmitter
Tracking transmitter
Command receiver and programming circuitry
Power package

Figure 1 is a schematic representation of th atmospheric structure satellite. The orbit wi be polar if we can launch into a polar orbit b the time the satellite is completed.

The location of the gages deserves speci attention. A pressure gage mounted on the forward or leading end of the satellite directl measures the density. When the gage is mounte on the side of the satellite the measured pressure is proportional to the ambient pressu itself. If the gage is at the back of the satelli the pressure provides a measure of the temperature, if the composition is known.

This satellite is spun, and as it goes throug orbit the orientation relative to the wind anges as a function of position around the rth. If the satellite is launched from north south in the middle latitudes the top inilly faces into the wind, but as it goes over e south pole and on around, the side faces to the wind. As it comes back around, the iling end is presented. With allowance for in, all side gages are presented at all aspects. With this device we hope to investigate furer the evidence obtained during the Intertional Geophysical Year for a change in deny with latitude, and for anomalous heating in e atmospheric thermosphere about the auroral nes.

Space probe—Table 2 lists representative inumentation to be included in a space probe.

TABLE 2—*Instrumentation for the space probe*
Basic scientific experiments

Cosmic rays and energetic particles (proportional counters)
Cosmic rays and energetic particles (scintillation counters)
Radiation (ion chamber and Geiger-Müller counter)
Magnetometer (search coil and optical aspect)
Micrometeorites
Temperature
Whistlers and VLF propagation
TV scanner
Engineering experiments
Telemeter package
olar power supply

is particular probe has been designed with phasis on the detection of particles in space. important feature of the instrumentation is capability for cross correlation. That is, asurements of the magnetic field are related the particle fluxes, and vice versa. The pare-detection package consists of proportional nters, scintillation counters, and an ion mber, the last measuring the total radiation. magnetometers include both the coil and flux-gate or saturable-core type.

lso included are a micrometeorite detector a temperature experiment, which have as r primary function the acquisition of data the space environment of the probe. Finally, TV scanner package will provide a low resoon image of the moon or any other bodies in range.

part of the instrumentation of the space probe must be concerned with technological or engineering problems. Telemetry and power supplies are the principal examples of engineering instrumentation.

Figure 2 is an artist's conception of the probe. In this particular device the solar cells are placed on paddles. The paddle geometry is designed to produce adequate power, after storage, to activate the instrumentation and the telemetering transmitter, regardless of the orientation of the probe axis relative to the sun. This probe might telemeter back information only a few minutes out of each hour, but we could track it possibly for several months or a year. The dimensions of the probe are somewhat less than 1 meter for the sphere and a span of 2 meters for the extended paddles.

Ionosphere satellite—Table 3 lists the equipment in a satellite that will make direct measurements in the ionosphere.

TABLE 3—*Instrumentation for the ionosphere satellite*

Two retarding potential chambers
Two ion traps
Two Langmuir probes
One electric field meter
7.75-Mc/sec antenna impedance probe
Command receiver
Tracking transmitter
Telemetering transmitter
Batteries
Despin mechanism

These instruments provide direct data, in contrast to the indirect propagation methods for ionospheric measurement, in which the effect of the ionosphere on signals transmitted from the rocket or satellite is observed. The electron density data are inferred from the propagation data by a number of techniques, perhaps the most useful being one in which two signals of different frequencies are transmitted from the object and compared at a ground receiver. The beat note between the two signals measures the electron densities in the region. In another technique used by Soviet scientists the times of radio rise and set of the satellite are observed. But in the present design for the ionosphere satellite the instrumentation will make direct local measurements on the electrons and ions in the immediate vicinity of the satellite. The instrumentation achieves this

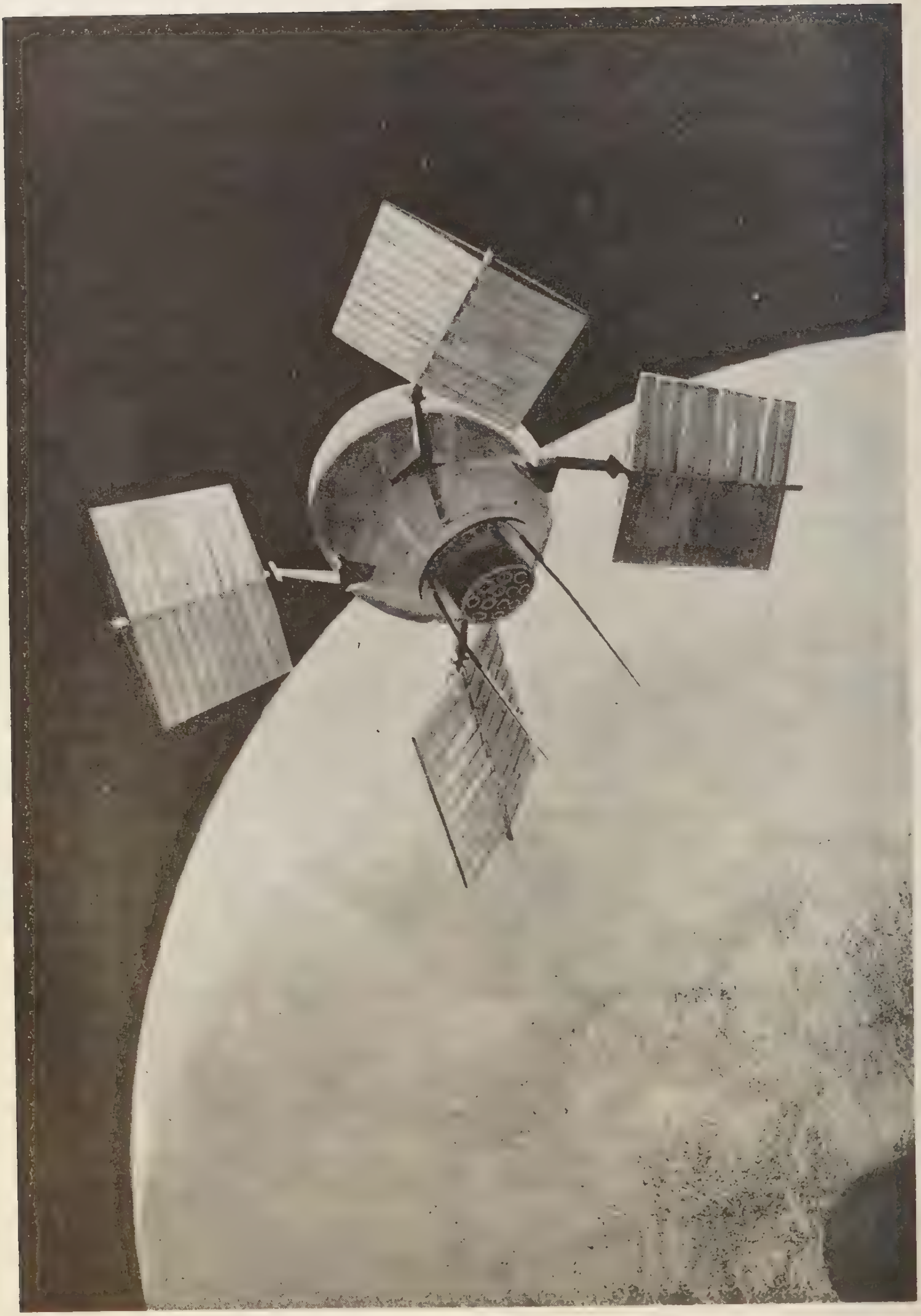


FIG. 2.—An Artist's conception of the space probe. The paddle-shaped devices contain the cells for solar power.

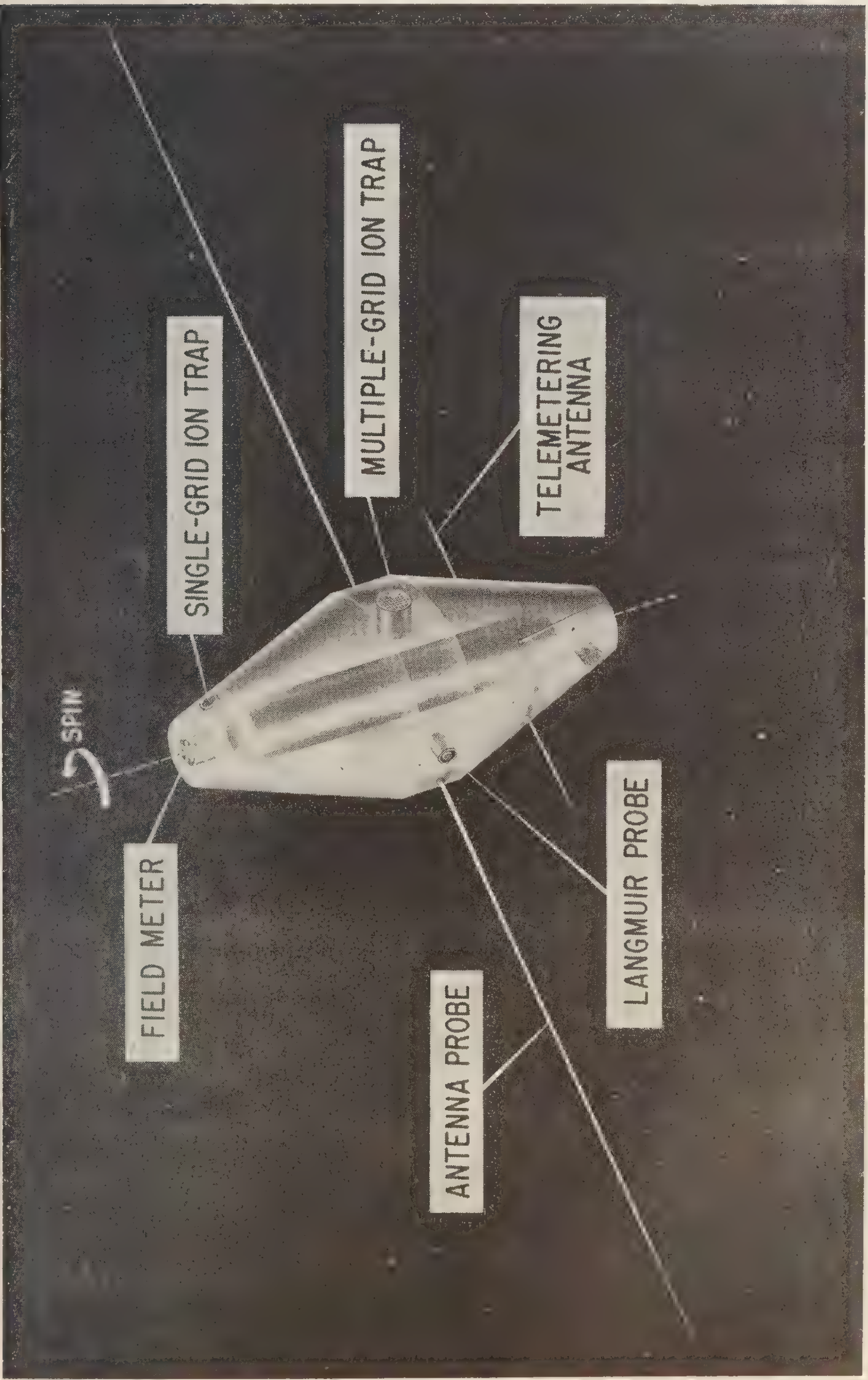


Fig. 3—A drawing of the ionosphere direct-measurements satellite; orbit, 135 to 800 miles; 50° inclination.

in two ways: there are Langmuir probes and ion traps to measure the electron and ion density directly; there is an antenna impedance probe to measure the effect of the surrounding plasma on the dielectric constant of the medium around two small antennas.

In greater detail, the instrumentation consists of two retarding potential chambers and two ion traps to obtain the concentration and ion spectrum, two Langmuir probes to determine the electron temperature, and one electric field meter to measure the charge of the satellite. This last instrument will check a recent suggestion that the satellite may acquire substantial charges in the radiation belts. The charge of the satellite is also necessary information for the interpretation of the ion and electron probe data. Finally there is the antenna impedance equipment which will measure the electron concentration by observing the frequency pulling of a low-powered transmitter feeding two small antennas.

Supporting instrumentation includes a command receiver, tracking and telemetering transmitters, batteries, and a despin mechanism, the last required because the launching vehicle will spin the satellite at too high a rate for proper measurements in orbit.

Figure 3 is an artist's conception of the ionosphere satellite. The retarding potential chamber is a simple two-element device which may be subject to photoelectric errors. The ion trap is more complicated, with four grids and a central collector. This is a more advanced ion probe, designed to reduce the disturbance of the medium and minimize errors associated with secondary emission effects. The present design is an advanced version of the ion trap used by the Russians on Sputnik III. The Langmuir probe is a single-element device of conventional design. The potential on the collector is programmed in a step function, and the current-voltage characteristics are telemetered to the ground. Figure 3 also shows the two antennas for the RF impedance experiment. A small transmitter feeds them. The dielectric constant of the medium will affect the antenna loading, producing a shift that can be telemetered to the ground and related to the electron concentration. This experiment has been successfully flown in a rocket.

The ionosphere satellite has dimensions of th order of a meter. The planned orbit will hav an inclination of 50°.

Orbiting astronomical observatory—Table lists the instrumentation in an astronomica satellite. The primary instrument in this sate lite is the X-ray spectrograph for solar radia tion. The others listed are those that will b compatible with it in size, weight, and functio Since X rays are known to affect the iono sphere, a VLF propagation experiment has bee included to provide data on the correlation be tween the intensity and spectrum of X rays an the structure of the *D* region. The planne spectrometer design uses a bent crystal for th region between 1 and 20 A and a grating fo the region between 20 and 100 A. The spectr graph will weigh 300 pounds and should have lifetime of 2 to 4 weeks for data transmissio

TABLE 4—*Instrumentation for the astronomical telescope*

X-ray spectrograph for solar radiation
VLF experiment to measure electron density
Other ion density and radio propagation experiments
Radiation measurements using broad-band optical filters and proportional ion chambers

Figure 4 is a drawing of the astronomic satellite. It indicates the first step in the deve opment of the stabilized space platform. Tl mounting of the X-ray spectrograph or oth device for solar spectroscopy will resemble t arm used on the 'pointing controls' develop for the Aerobee rocket and discussed by I Friedman in the preceding paper. The great part of the mass of the satellite is on the perir eter of the disk. The spin of the disk is mai tained by two jets fed from a helium-fill sphere in the center of the satellite. The je operate only when the spin rate falls below critical value. The satellite contains two otl jets, fed by the same helium supply, whi control the orientation of the symmetry a of the satellite.

Magnetometer space probe—Table 5 describ the last object, designed for magnetic fi measurements from the earth to the vicinity the moon and beyond. The primary instrum in this probe is a rubidium-vapor magneto eter, shown mounted in the probe in Figure The operation of the magnetometer require

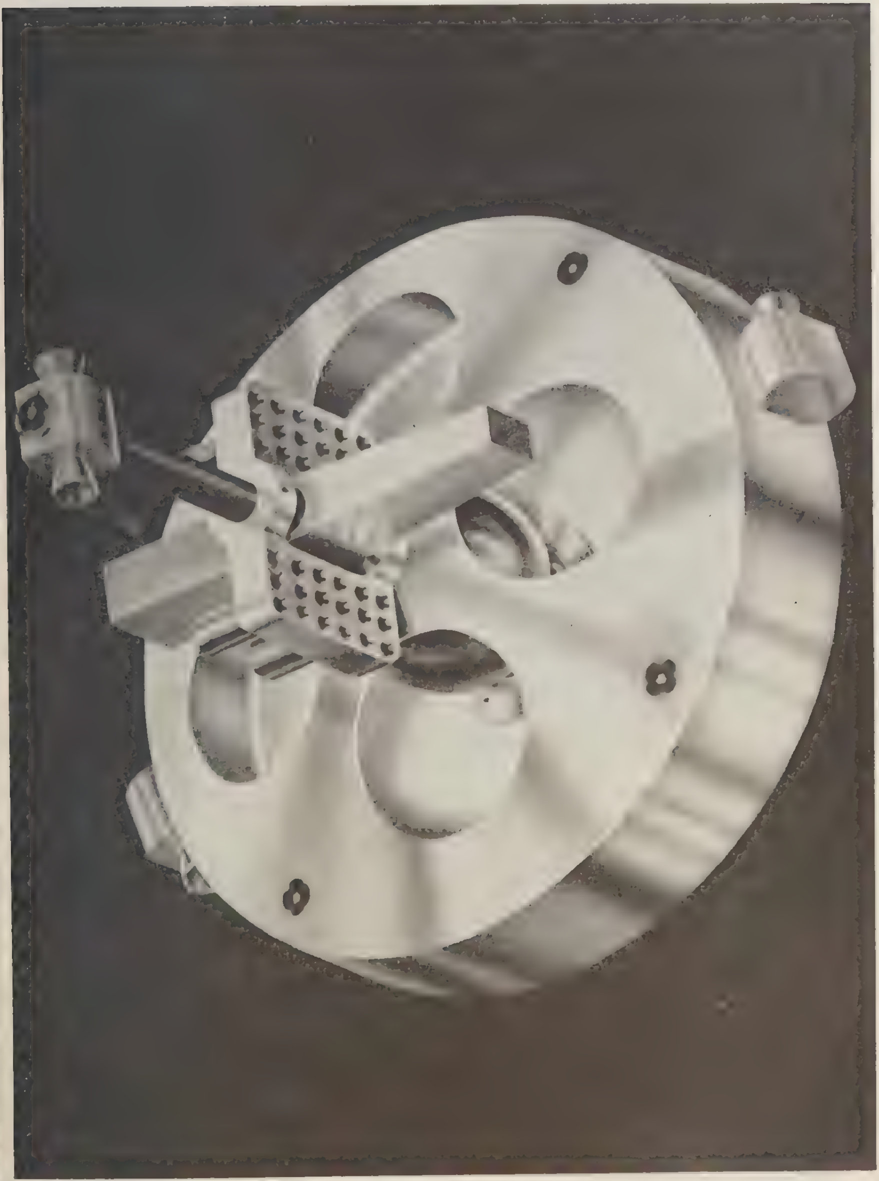


FIG. 4—A drawing of the astronomical satellite.

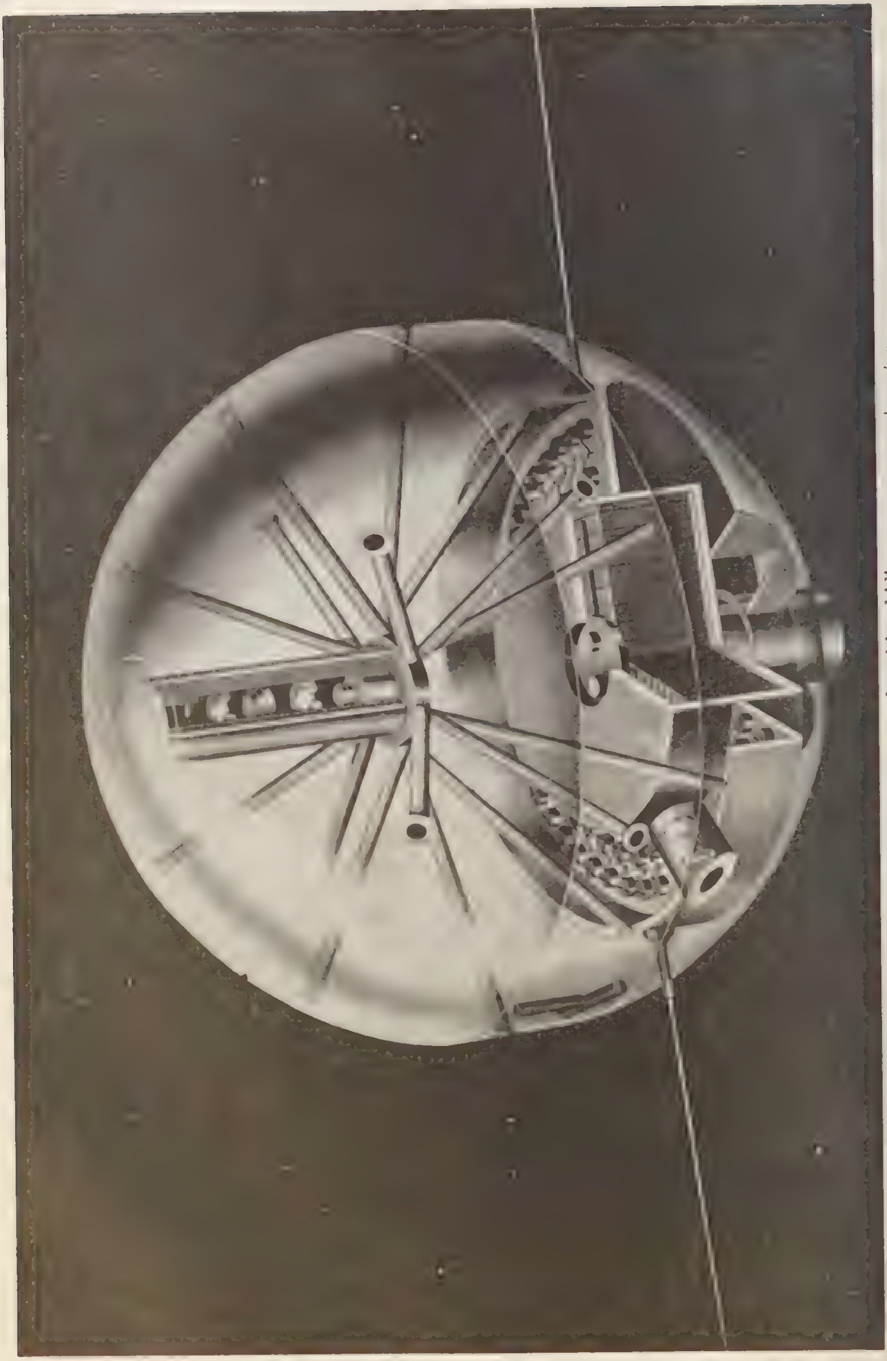


FIG. 5—A drawing of the space probe with rubidium-vapor magnetometer.

Table 5—*Instrumentation for the magnetometer probe*

Rubidium-vapor magnetometer
Two flux-gate magnetometers
Optical-aspect sensors
Telemetry transmitters
Duty-cycle programmer
Batteries

source, in this device an electrodeless discharge, which excites the red line of rubidium at 7000 A. The light from this source is polarized and passes through an absorption cell, and then through a lens system to a photocell. The photocell output is connected to a feedback circuit going back into the absorption cell. The feedback circuitry is designed to maintain the absorption at a maximum in the cell. The frequency in the feedback loop is then a function of the absolute value of the magnetic field. The basic physical constants are such that in a field of 1 gauss the natural frequency is 700 kc/sec. The threshhold sensitivity of the instrument is a fraction of a gamma. Figure 5 shows the electronics necessary to operate the magnetometer. Liquid batteries are in the center, and two flux-gate magnetometers are placed along the edge for a measurement of the vector field.

The three satellites and two probes just described are second-generation experiments. Some of them are under construction now. I believe that they will be representative of devices in the intermediate-weight class. Before the last of them is fired a number of the still larger payloads will have been launched. But probes and satellites in this class will be a continuing part of our space science program, because they will always provide the opportunity for exploratory experiments at a minimum cost.

Discussion

Question: In regard to the later larger astronomical telescopes, are you giving consideration to the actual recovery of the film in a nose cone? Some of the people in nose-cone technology think one could recover this from a satellite, detach part of it, and actually find where it hits the earth.

Mr. Townsend: In the case of energetic particle detectors, we would like very much to get back emulsions from very high altitudes, and we are working on recovery schemes for this purpose. There are some methods available now, but we are working on others that are less expensive and will return relatively small packages, perhaps 10 pounds, from sounding rockets that go out several thousand miles. If these techniques are successful, I think that a family of recovery devices will develop for sounding rockets. And then, possibly as a by-product of other large programs, such as Project Mercury, the man-in-space program, we may have means for recovering very large devices containing spectrographs and astronomical equipment.

Chairman: This paper is particularly important to our symposium, because it tells those of us who are not actively working on the construction of space apparatus a great deal about what can be done, and how it must be done.

Outer Atmospheres of the Earth and Planets

ROBERT JASTROW

National Aeronautics and Space Administration
Washington, D. C.

Introduction—We shall confine our attention those major objects in the solar system that e readily accessible, namely, Mars, Venus, the oon, and the earth. We hope that our present servational knowledge of the properties of the st three of these bodies will be replaced within e next decade by detailed information based direct observation. The prospect of lunar d planetary exploration by space vehicles is very exciting one, both for the scientific mmunity and for the general public. A subantial part of the NASA effort in space rearch will be channeled into these investigaons.

For the present we must be satisfied with the agmentary information obtainable by terstrial observation. Very little is known about e atmospheres of Mars and Venus in parcular, and we can summarize the salient facts out these two planets quite briefly in the st paragraphs of our discussion. We shall en discuss the lunar atmosphere at somewhat eater length, and conclude with a presentation recent developments in the physics of the rth's outer atmosphere.

Mars and Venus—The development of the artian atmosphere probably followed that on e earth, with allowance for variations in the al stages produced by lower temperatures d a reduced value of surface gravity.

The present abundance of O_2 in the Martian mosphere is expected to be substantially less an that on the earth for several reasons. First, e internal temperatures are lower and the oduction of steam correspondingly less. Secd, the temperature minimum at the tropouse is believed to be lower than on the earth; is implies a reduction in the water content ove the altitude of the tropopause and hence reduction in the rate of production of O_2 by otodissociation [*Kuiper*, 1952].

The result of a search for the O_2 line in the Martian spectrum suggests an upper limit of 10^{-3} for the ratio of Martian to terrestrial O^2 abundance, as might be expected from the above arguments.

Kuiper has shown that CO_2 occurs on Mars with a somewhat greater abundance than in the terrestrial atmosphere. H_2O probably exists in small quantities in the form of ice or frost extending over the polar regions. Argon should be present as the decay product of radioactive potassium.

CO_2 is the principal constituent of the Venusian atmosphere, occurring with an abundance far in excess of that on the earth. According to an argument by Urey the abundance of CO_2 implies the absence of water in the present atmosphere of Venus [*Urey*, 1951]. Urey remarks that in the terrestrial planets atmospheric CO_2 will be depleted by the weathering of rocks according to the reaction

$$MgSiO_3 + CO_2 \rightleftharpoons MgCO_3 + SiO_2$$

This reaction takes place in the presence of H_2O, and only then. Therefore the large CO_2 content in the atmosphere of Venus implies that water does not exist at the present time above the surface of that planet. It should be noted, however, that the surface may be completely covered with water to a substantial depth; in that case the weathering of rocks will cease when a layer of carbonate is formed, unless underwater currents are strong enough to produce appreciable erosion [*Menzel and Whipple*, 1955].

At an earlier stage in the development of Venus H_2O must have been present, since the abundance of CO_2 is presumably due to the production of O_2 derived from H_2O by photodissociation.

Argon produced by the decay of radioactive potassium should be present on Venus in the same abundance as on the earth. Some N_2 should also be present as a product of volcanism.

The moon—The composition and extent of the lunar atmosphere are determined by the competition between the processes of accumulation of gases from the surface and escape from the atmosphere of those molecules with exceptionally high velocities—sufficient to remove them from the gravitational attraction of the moon.

Probable sources for the accumulation of surface gases are: the radioactive decay of potassium, producing argon; and residual volcanic activity, producing such gases as SO_2, CO_2, and H_2O.

Estimates of the emission of these gases based on terrestrial data suggest a source strength of 5×10^6/cm^2/sec for argon, and 10^{10}/cm^2/sec for the volcanic gases [*Vestine*, 1958].

The moon has 1/80 the mass and ¼ the radius of the earth. The gravitational potential is therefore 1/20 that at the surface of the earth. The reduced gravitational potential indicates that gases will escape much more rapidly from the lunar than from the earth's atmosphere.

The rate of escape also depends on the maximum temperature, which determines the number of atoms or molecules in the high-energy tail of the thermal distribution. Thermocouple measurements indicate a value of 370°K for this temperature [*Wesselink*, 1948].

Substituting the above values of J and T into Maxwell distribution formulas, we find for argon $n_0 \simeq 5 \times 10^{15}$/cm^3, or 10^{-4} atmosphere. A similar calculation for the volcanic gases leads to pressures greater than 1 atmosphere for SO_2 and CO_2, and to 4×10^{-8} atmosphere for H_2O. The latter values will be lowered substantially, however, when allowance is made for the effects of photodissociation or chemical reactions with the crust, and for the fact that the age of the moon ($\simeq 4.5 \times 10^{-9}$ yr) does not allow enough time for SO_2 and CO_2 to build up to equilibrium [*Vestine*, 1958].

The recent calculations by *Herring and Licht* [1959] have shown that these estimates of the extent of the lunar atmosphere are seriously reduced by the effects of the solar wind. It appears, in fact, that the solar wind will blow away all but a very small fraction of the lunar atmosphere estimated on the basis of conventional calculations of the probability of escape

Herring and Licht consider a solar wind consisting of protons with the conventional densit of 10^3/cm^3, and velocity of 10^8/cm/sec, or energies of 10 kv [*Biermann,* 1951]. In an elasti collision with an atom these protons will transfer an average of 1 kev of kinetic energy, whicl is sufficient for the escape of the atom. If w assume that every atom struck by a proto escapes, the rate of injection of particles i readily calculated.

For argon, Herring and Licht find an equilibrium density of 10^4 particles/cm^3, or 10^{-} atmosphere, remaining when the solar wind i taken into account. For volcanic gases they fin a density of 10^8 particles/cm^3, or 10^{-11} atmosphere. This last value represents an upper limi for the volcanic gases because photodissociatio has been neglected.

Comparing these values with the results obtained above, we see that the solar wind reduces the density of the lunar atmosphere by factor of 10^{11} for argon and 10^{15} for CO_2.

The results quoted by Herring and Licht depend on the assumption that the moon's magnetic field is too feeble to shield the atmosphe effectively from the solar wind. Assuming a upper limit of 200 gamma for the selenomagnet field [*Clement*, 1956], we find that the magnet pressure of a field of this intensity will indeed have no shielding effect.

The values quoted for argon and the volcar gases lie to either side of an experimental estimate of 10^{-13} atmosphere, which has been derived from measurements of the lunar ionosphere by *Elsmore and Whitfield* [1955].

Temperature and density in the terrestr atmosphere—During the course of the International Geophysical Year the rocket and satell programs of the United States and the Sov Union yielded a substantial amount of data the properties of the ionosphere and the densi and temperature of the upper atmosphere. T analysis of these measurements has reveal several features which promise to broaden c understanding of the influence of solar corpuscular streams on upper atmosphere properti These developments rank with the most interesting results of the IGY upper-atmosph program.

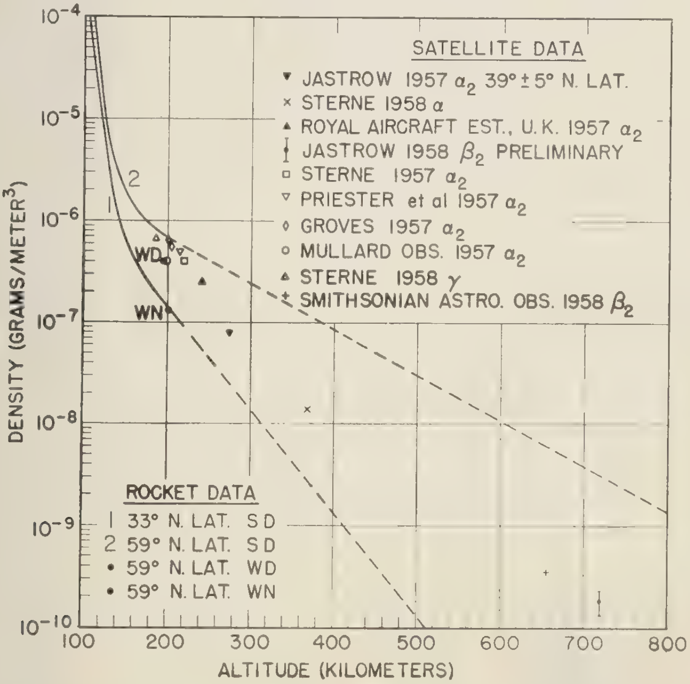
Analysis of satellite data—At the beginning ıe IGY the density and temperature of the upper atmosphere were relatively well determined elow 100 km, and known with less precision up › 200 km [*Newell,* 1955; *Kallmann and others,* 956]. Above 200 km the properties of the atıosphere could only be estimated by extrapolaons from the data at lower altitudes [*Minzner ıd Ripley,* 1956]. These estimates were uncerıin by a factor of 100 at an altitude of 500 km.

The ceiling on density measurements was lifted › 700 km by the first data from the Vanguard satellite, which yielded values at that altitude ith a factor of uncertainty of 2. Other sateles before and after Vanguard I filled in the ensity curve at intermediate altitudes with the ıme precision [*Schilling and Sterne,* 1959; *arris and Jastrow,* 1959; *King-Hele,* 1959].

The density of the atmosphere is determined om satellite tracking data by an indirect ethod, based on the analysis of changes in the riod of revolution of the satellite. The change in the orbital period is a direct result of atmospheric drag, which causes the satellite to lose energy continuously during its lifetime. As the energy of the satellite decreases, it falls toward the center of the earth, increasing its velocity and reducing its average altitude, and therefore reducing the time required to complete each circuit. Detailed calculations, based on the equations of satellite motion, then determine the quantitative relation between the reduction in the period and the average air density in the orbit. Figure 1 shows density values obtained in this way from the analysis of the orbits of a number of United States and Soviet satellites [*LaGow and others,* 1958].

Since the density falls off very rapidly with increasing altitude, satellite density determinations are heavily weighted by the contributions from perigee, the lowest point in the orbit.

Satellite drag fluctuations—The satellite density data show large fluctuations over periods of



ɪɢ. 1.—Compilation of rocket and satellite measurements of upper atmosphere densities. The rocket ɑa at 33°N and 59°N refer to White Sands and Fort Churchill, respectively. *WD, WN,* and *SD* resent winter day, winter night, and summer day conditions, respectively, at Fort Churchill. rves 1 and 2 are drawn through rocket data [*LaGow and others,* 1958].

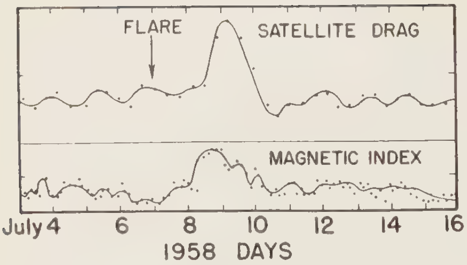


FIG. 2—Correlation of satellite drag variations with magnetic storms at the time of a major solar flare [*Jacchia, 1959*]. The occurrence of the flare is indicated by the arrow. A comparison of the upper and lower curves indicates that the drag increase and the magnetic storm both began approximately 36 hours after the flare.

a few days or weeks, when the accuracy of the tracking data is sufficient to permit the determination or orbital changes in such relatively short intervals [*Jacchia*, 1959; *King-Hele and Walker*, 1959]. Figure 2 shows an example of the time variations in drag, calculated by Jacchia from the optical data on Sputnik III. These variations are probably the result of atmospheric heating produced by the incidence of streams of energetic particles and radiation from the sun. The surface of the sun boils and bubbles in a very active manner, occasionally emitting large gusts of plasma and radiation into the solar system. It has been discovered by Jacchia that the apparently random fluctuations are actually proportional to the changes in the intensity of the 10-cm radiation from the sun, which constitutes an excellent measure of solar surface activity. Furthermore, the fluctuations show a tendency to repeat every 27 days, which is the period of rotation of the sun about its axis. Thus it appears that in addition to the heating and expansion of the atmosphere produced by steady solar exposure, there is further heating caused by streams from definite spots on the surface of the sun, which come around every 27 days in the course of the sun's rotation.

In certain cases, those following major solar flares, the increase in drag acting on Sputnik III has been tentatively identified with the arrival of corpuscular streams in the vicinity of the earth [*Jacchia*, 1959a]. Figure 2 shows that the drag increase on Sputnik III observed at the time of a major solar flare did not occur simultaneously with the flare, but began approximately one day later at the same time that th magnetic K index spurted upward. A rise i the K index signifies the onset of a magneti storm, and therefore the arrival in the aurora region of the relatively slow-moving solar corpuscular stream which accompanies the flare These solar particles are directed by the geomagnetic field into auroral latitudes where the transfer energy to the atmosphere by mechanisms such as that discussed below. Sputnik II passes through the auroral zone and was apparently affected by the heating and expansio of the atmosphere in that region.

The rocket data—Density measurements ar also made directly by ionization gages installe in rockets. The most interesting characterist of the rocket data is their strong latitude dependence. The experiments, carried out by LaGow, Townsend, and their collaborators at th Naval Research Laboratory, indicate that 200 km the summer daytime density in arct latitudes is about 6 times greater than the corresponding density in temperate latitudes [*LaGow and others*, 1958; *Townsend and Meadow* 1958]. LaGow notes that this latitude differen may not be representative, however, since th flight in the temperate zone occurred in 19 during the sunspot minimum, whereas the arct flight took place at the peak of the sunsp activity in 1957.

It is also found that the density values in th arctic zone depend on the season and the tin of day. The data show that at an altitude 200 km the northern atmosphere is 2 tim heavier in the day than at night, and 2 tim heavier in the summer than in winter. The results are shown in Figure 1.

LaGow has suggested a reasonable explanati for the diurnal, seasonal, and latitudinal variations. During the day, exposure to the sun probably heats the atmosphere and causes an upward expansion, producing a large relative increase in the density of the thin air at hi altitudes. The effect of solar exposure shou be greater in summer than in winter, and greatest of all during the long day of the arctic summer when the exposure to the sun is near continuous. These predictions are in agreeme with LaGow's observations. The evidence presented below suggests, however, that solar e posure may not provide the complete explanation of the observed latitude variations.

Temperature—Although direct temperature measurements have not been made in the upper atmosphere, the temperature can be determined indirectly from the altitude dependence of the density measurements with the aid of the formula

$$d = d_0 e^{-(h-h_0)/H} \tag{1}$$

which gives the relation between the density d at altitude h and density d_0 at altitude h_0. The scale height, H, is obtained from the formula $H = kT/\bar{M}g$, in which k is the Boltzmann constant, T and $\bar{M}$ are, respectively, the temperature and molecular weight of the atmosphere, and g is the acceleration of gravity. This result is valid provided that the collision mean free path in the atmosphere is less than H and that the fractional change in temperature is small over a distance of H.

According to equation 1, the temperature of the air is obtained directly from the slope of a logarithmic plot of density vs. altitude. By this method LaGow and his collaborators determined atmospheric temperatures at altitudes between 150 and 200 km over the White Sands missile range in New Mexico, and over the Fort Churchill range in the Canadian auroral zone. They found that during a summer day the temperature of the air in the auroral zone is approximately 2200°K, as compared with a relatively cool 1100°K over New Mexico [*Horowitz and LaGow*, 1957, 1958]. Figure 3 shows their results joined to earlier measurements of atmospheric temperatures.

Rocket and satellite disagreement—The satellite results do not show the latitude dependence that appears in the rocket data. Satellite measurements always give the density at the position of perigee, and a satellite therefore automatically samples a broad range of latitudes during the course of the rotation of its perigee in the plane of the orbit. For the Vanguard I satellite, for example, the latitude of perigee changes from 33°N to 33°S every 41 days. Although Vanguard I has been transmitting signals for more than a year, the orbit data have shown no significant changes thus far that can be correlated with latitude variations. The contrast with the strong latitude effect in the rocket data constitutes an outstanding puzzle in the interpretation of the IGY results. The attempt to understand this inconsistency leads us to some of the most interesting implications in the density data.

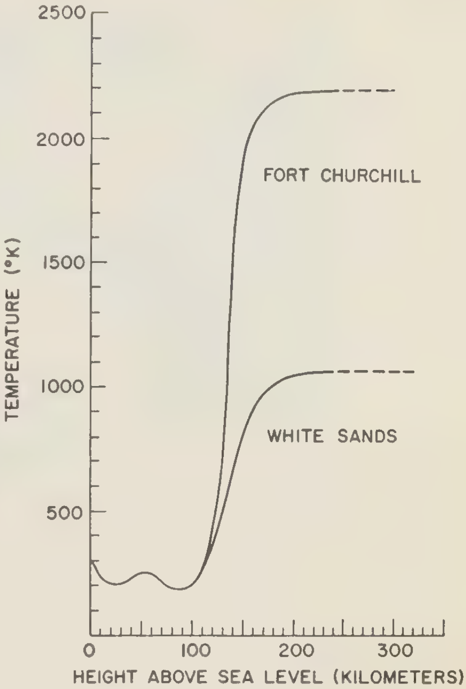


Fig. 3—Rocket measurements of upper atmosphere temperatures. The White Sands data refer to a latitude of 33°N; the Fort Churchill data were obtained at latitude 59°N in the auroral zone [*Horowitz and LaGow*, 1958].

Effect of the Van Allen belt—The particles in the Van Allen belt may provide the explanation for the difference between densities measured at White Sands and at Fort Churchill. The Van Allen particles are trapped by the geomagnetic field in orbits in which they spiral along the lines of magnetic force. As Figure 4 indicates, the particles in the outer belt are funneled into the arctic and antarctic zones by the concentration of the magnetic field near the north and south poles. The outer belt dips down into the atmosphere in these regions and disturbs the normal conditions that exist at lower latitudes.

The interaction between the trapped particles and the atmosphere produces two major geo-

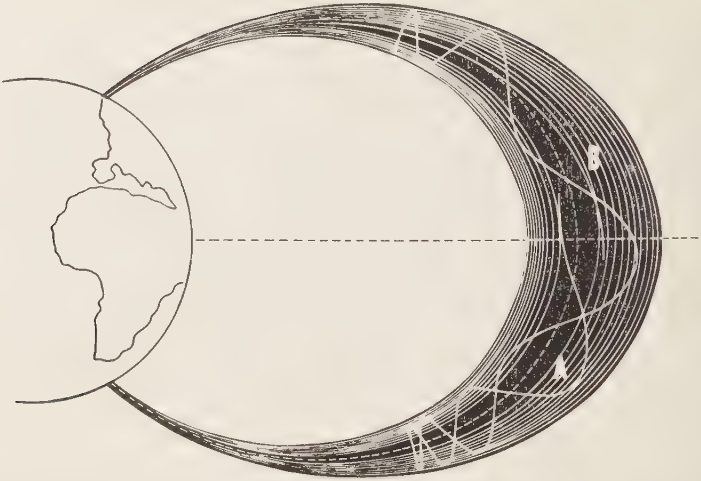


Fig. 4—Channeling of geomagnetically trapped particles into the auroral regions. Curves A and B, taken from the calculations by *Störmer* [1913], show the trajectories of particles injected into the outer zone at large and small angles, respectively, to the direction of the local magnetic field.

physical effects. First, the temperature of the auroral zone is raised by the energy transferred in collisions between the atoms and molecules of the upper atmosphere and the trapped particles in the outer belt. Second, the aurora borealis and the aurora australis may result from the excitation of the arctic and antarctic atmospheres by these same collisions with the particles trapped in the outer zone.

Theories on the heating of the upper atmosphere by the channeling of charged particles in the geomagnetic field predate the discovery of the Van Allen belts [*LaGow and others*, 1958]. Van Allen, McIlwain, and Ludwig returned to this possibility in their first discussion of the Explorer IV data, in which they suggested that the trapped particles may produce both auroral excitations and atmospheric heating [*Van Allen and others* 1959]. Soviet scientists have made similar suggestions in their interpretation of the results from Sputnik III [*Krassovsky*, 1958].

The Van Allen data, supplemented by Sputnik III results and by rocket measurements performed by the Naval Research Laboratory and the State University of Iowa, now provide the first opportunity for the theorist to make a quantitative test of these interesting ideas. In order to carry out the necessary calculations, we have constructed an idealized model in which the heating effect of the Van Allen layer is confined to a definite region in the auroral zone whose boundaries are kept at a 'temperate' level of 1000°K. The rate of heat transfer to the air within this zone is calculated from the combination of all rocket and satellite observations on the Van Allen particles. Figure 5 is a schematization of the model. The region labeled 'outer zone' in the figure indicates a volume within which we assume a heat source of Q cal/cm³/sec given by

$$Q = F\sigma\bar{E}n \qquad (2)$$

where

F = flux of energetic electrons.
σ = cross section for inelastic collisions.
$\bar{E}$ = energy transferred per collision.
n = number density of atmospheric particles

The flux of Van Allen particles must be taken from the Sputnik III data, since the Explorer IV trajectory does not penetrate into the center of the auroral zone. From the Sputnik III results we have at present only one datum, namely, a reported energy flux of 4000 ergs/cm²/sec for electrons above 10 kev [*Krassovsky*, 1958]. This value refers to an altitude over Russia which may be estimated as 300 km from our knowledge of the Sputnik III orbit. At 300 km, $n \approx 2.5 \times$

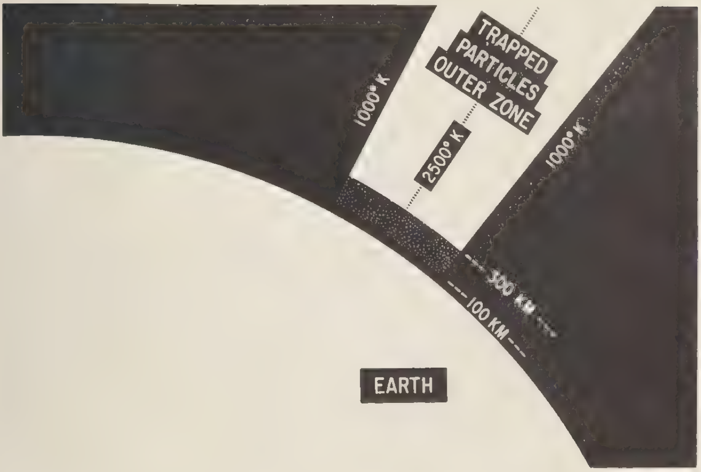


Fig. 5—Schematization of the model used for the calculation of atmosphere heating by energetic electrons.

10^{-9}/cm^3. Finally, at 10 kev the cross section (σ) for inelastic collisions is approximately 10^{-17} cm^2, and the kinetic energy ($\bar{E}$) available in each collision is $\approx$20 ev (Fite, private communication, 1959).

Using the values given above, we find $Q \approx 4 \times 10^{-16}$ cal/cm^3/sec at 300 km. We assume that the trapped particle flux varies in inverse proportion to the number density for atmospheric particles. On this assumption, the heat source is independent of altitude.

Recent rocket observations suggest an additional source of excitation at the time of auroral displays. The rocket experiments were performed by *Meredith and others* [1958] of the Naval Research Laboratory and *McIlwain* [1958] of the State University of Iowa. They indicate that in the auroral zone and at the time of an aurora there is a substantial flux of energetic electrons with roughly constant intensity between 100 and 180 km. These energetic electrons have a uniform angular distribution over the upper hemisphere. In this respect they differ from the trapped Van Allen particles, whose angular distribution is concentrated in the plane perpendicular to the local direction of the magnetic field. The isotropic distribution of the electrons observed in the rocket experiments suggests that they are not trapped electrons, but rather the electrons that have been removed from the trapped particle layer by multiple scattering and now wander down through the atmosphere beneath the lower border of the trapped-particle zone. The flux (F) of this group of electrons is observed to be roughly independent of altitude. According to equation (2) their rate of energy transfer therefore varies only with atmospheric density (n). Hence the corresponding heat source is a sharply increasing function at lower altitudes. The superposition of this source on the altitude-independent source produced by the trapped electrons leads to the curve at the right-hand side of Figure 6.

The isotropic electrons do not make a significant contribution to the calculated temperature increase. We have added them to the heat source because of their relevance to the problem of auroral excitation, which is discussed below.

The equation of heat conduction, $Q = K\nabla^2T$, may be solved analytically with the indicated boundary conditions. Harris and Jastrow have shown in this way that the temperature must rise to approximately 2500°K at the center of the auroral zone. The close agreement with observation is fortuitous in view of the approximations made in Q and in the construction of the boundary value problem. It is significant, however, that the temperature increase has the correct order of magnitude.

Atmospheric models—The preceding analysis

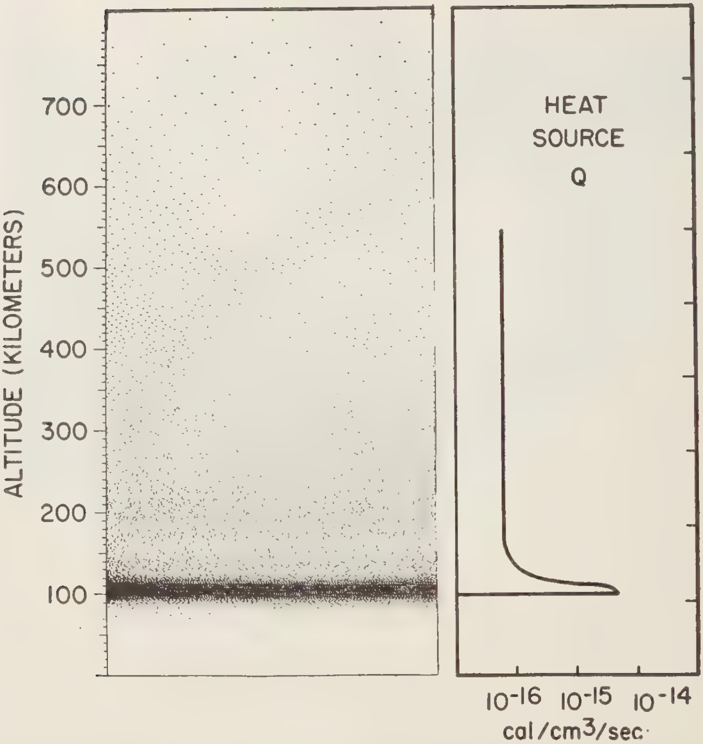


FIG. 6—Comparison between the distribution of auroral altitudes (*left*) and the calculated rate of energy transfer from energetic electrons to the atmosphere (*right*).

suggests a model of the upper atmosphere in which densities in the auroral zone are strongly affected by corpuscular heating. If this picture is correct we must consider two distinct atmospheric systems, one for the auroral zone and the other for lower latitudes.

The first model, which may be called the auroral atmosphere, is represented by curve *A* in Figure 7. The auroral atmosphere consists of rocket observations at Fort Churchill extrapolated above 200 km on the assumption of a constant temperature. The assumption of constant temperature is based on the LaGow measurements which indicated little temperature variation between 170 and 210 km.

The second model, which may be called the temperate-equatorial atmosphere, is shown as curve *T* — *E* in Figure 7. It is obtained by drawing a smooth curve through the White Sands rocket data and the data obtained from the Explorer and Vanguard satellites. All density data used in the construction of this model were obtained at latitudes no greater than 33°.

Origin of the aurora—Since the energy content in the Van Allen belt may be sufficient to make substantial changes in the temperature and density of the upper atmosphere and perhaps to account for the observed latitude effect in the rocket data, we are encouraged to look further into the possibility that the outer Van Allen belt also provides the origin for auroral phenomena. As a first step, I. Harris and the author examined the data on the frequency of auroral events as a function of altitude, which have been collected and published by *Störmer* [1955]. We presumed that, if the energetic Van Allen electrons are the primary source of the aurora, the rate of energy transfer from the Van Allen particles to the atmosphere will also govern the altitude distribution of auroral displays. In Fig-

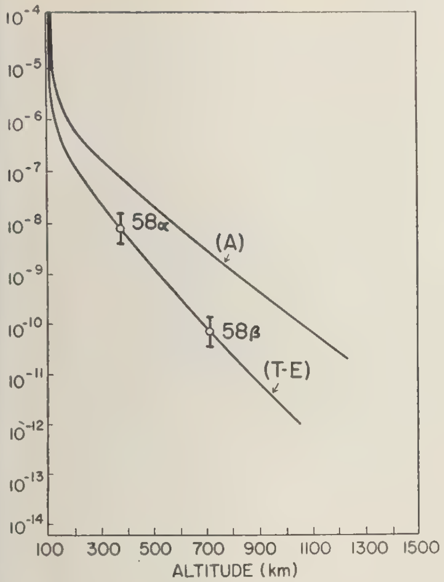


FIG. 7—Density of the upper atmosphere. The — *E* curve describes a temperate-equatorial ɪmosphere, based on rocket and satellite data ɔtained at latitudes less than 33°; curve *A* ·fers to an auroral atmosphere obtained by ktrapolation of the Fort Churchill rocket data ɔove 200 km on the assumption of constant ·ale height [*LaGow and others*, 1958]. The satelte points are derived from the Explorer I l958α) and Vanguard I (1958β) orbits [*Harris ɪd Jastrow*, 1959].

re 6 we show on the right the heat source *Q* ι calories per cubic centimeter per second estiιated from the rocket and satellite data on the ɪxes of energetic electrons, as determined in ɪr studies of temperature variations in the pper atmosphere. On the left in Figure 6 we ιow the altitude distribution of 12,330 auroras, ; reported by Störmer. We see that the intenty of electron energy transfer to the atmoshere and the frequency of auroral events both sappear below 90 km and have a sharp maxiɪum near 100 km. In our view this correspondɪce strongly supports the suggested association ɛtween energetic electrons and auroral disays. Additional evidence of a very convincing ɪture is supplied by the recent analysis of *estine and Sibley* [1959] in which it is shown ɪat the theoretical properties of the trapped electron layer lead to an accurate prediction of auroral isochasms in the arctic and antarctic zones.

References

BIERMANN, L., *Z. Astrophys.*, *29*, 274, 1951.
CLEMENT, G. H., Rand Corp., Paper P-833, rev. May 7, 1956.
ELSMORE, B., AND G. R. WHITFIELD, *Nature*, *176*, 457, 1955.
HARRIS, I., AND R. JASTROW, *Planetary and Space Sciences*, *1*, 20, 1959.
HERRING, J. R., AND A. L. LICHT, *Science*, *130*, 266, 1959.
HOROWITZ, R. AND H. E. LAGOW, *J. Geophys. Research*, *62*, 57, 1957.
HOROWITZ, R. AND H. E. LAGOW, *J. Geophys. Research*, *63*, 757, 1958.
JACCHIA, L. G., *Nature*, *183*, 526, 1959.
JACCHIA, L. G., *Nature*, *183*, 1662, 1959a.
KALLMANN, H. K., W. B. WHITE, AND H. E. NEWELL, *J. Geophys. Research*, *61*, 513, 1956.
KING-HELE, D. G., *Nature*, *183*, 1224, 1959.
KING-HELE, D. G., AND D. M. C. WALKER, *Nature 183*, 527, 1959.
KRASSOVSKY, V. I., presented at the Fifth CSAGI Assembly, Moscow, July, 1958.
KUIPER, G. P., *Atmospheres of the Earth and Planets*, chapter XII, University of Chicago Press, 1952.
LAGOW, H. E., R. HOROWITZ, AND J. AINSWORTH, *IGY Rocket Rept. Series*, *1*, 38, 1958.
MCILWAIN, C. E., *IGY Rocket Rept. Series*, *1*, 164, 1958.
MENZEL, D. H., AND F. L. WHIPPLE, *Pubs. Astron. Soc. Pacific*, *67*, 161, 1955.
MEREDITH, L. H., L. R. DAVIS, J. P. HEPPNER, AND O. E. BERG, *IGY Rocket Rept. Series*, *1*, 169, 1958.
MINZNER, R. A., AND W. S. RIPLEY, *AF Surveys Geophys.*, *86*, 201, 1956.
NEWELL, H. E., *Ann. Géophys.*, *11*, 115, 1955.
SCHILLING, G. F., AND T. E. STERNE, *J. Geophys. Research*, *64*, 1, 1959.
STÖRMER, C., *Videnskapsselskapets-Skrifter*, *14*, 1913.
STÖRMER, C., *Polar Aurora*, Oxford University Press, 1955.
TOWSEND, J. W., AND E. B. MEADOWS, *IGY Rocket Rept. Series*, *1*, 14, 1958.
UREY, H. C., *Geochim. et. Cosmochim. Acta*, *1*, 209, 1951.
VAN ALLEN, J. A., CARL E. MCILWAIN, AND GEORGE H. LUDWIG, *J. Geophys. Research*, *64*, 271, 1959.
VESTINE, E. H., AND W. L. SIBLEY, Rand Corp. P-1726-NSF, 1959.
VESTINE, E. H., Rand Corp., RM-2106, 1958.
WESSELINK, A. J., *Bull. Atron. Inst. Netherlands*, *10*, 351, 1948.

Discussion

Question: I should like to make a comment about the lunar atmosphere. Radio observations make clear that there are no more than a few thousand electrons in excess of the background density in the solar system in the vicinity of the moon. That is to say, if you place any amount of gas there, then of course you expect, if it is permanently there, the generation of an ionosphere, and you would expect electron densities that much exceed this. Therefore, this places a very low limit on the amount of gas that you would expect, quite apart from the optical measurements.

I think the solar wind is a dominating influence for removing any gas, and I can really see no way for a planet to keep an atmosphere unless it can protect itself with some hydrogen. It has to be able to hold hydrogen for at least a short time in order that the solar wind shall avoid this sweeping away of the other gases.

The atmosphere that one would expect a body like the moon to have is, however, the temporary effect, which is quite a substantial one, due to gas coming in from the sun, frequently or always —we are not sure—at speeds of the order of a thousand kilometers per second, with particle densities of the order of a few hundred, and leaving in a steady state again from the moon at speed that cannot be much in excess of the thermal speeds, because the particles penetrating onto the moon will evaporate again at something of the order of thermal speeds, which are a few hundred times slower. Therefore, continuity demands an increase of density by a factor of a few hundred over the solar wind density.

This would be an atmosphere that would be constantly blowing off the moon, and be swe back like the comet phenomenon. A small fractio of that, but a very small fraction of that, in tur will get ionized fast enough—because it will hav been deionized on vaporizing off the moon—w get ionized fast enough so as to make a few ext ions on the way out. Most of the stuff sweepi out will be neutral.

Mr. Jastrow: With regard to the ionizatio level, the upper limit on the electron density ma well be a real one, but the percentage of ioniz tion in the lunar atmosphere is a very difficu matter to estimate, and I would expect it to b small for the following reason: Let us think of th argon atmosphere, for example, for which th estimates yield a total of 100 tons, and of th very low densities which are implied by ou estimates. At these densities a typical argon ato spends several hundred seconds in the air befo it returns to the surface. It leaves the surfa neutral, and our estimates of photoionizatio cross sections indicate that the times for ioniz tion are much longer than this few hundre seconds. So one expects a very small ionizatio percentage.

Question: I should like to ask if an investig tion has been made of the consequences of the high auroral temperatures on the escape of th earth's atmosphere.

Mr. Jastrow: No, but one might ask whethe when one gets up to the base of the exosphe the temperature differences between this auror region and the temperate equatorial atmospher are appreciable. One might guess that tho differences have largely disappeared at that alt tude. We were, after all, talking about altitud of two or three hundred kilometers.

Round-Table Discussion

Chairman: Lyman Spitzer, Jr.

Princeton University Observatory
Princeton, New Jersey

Participants: H. Friedman, S. Fritz, L. Goldberg, R. Jastrow, D. H. Menzel, H. E. Newell, A. H. Shapley, J. W. Townsend, Jr., H. C. Urey

Mr. Spitzer: Let me introduce the members of the panel. Starting at this end: Mr. Goldberg, University of Michigan Observatory; Mr. Menzel, Harvard University; Mr. Newell, NASA; Mr. Shapley, Bureau of Standards, Boulder, Colorado; Mr. Jastrow, NASA; Mr. Townsend, NASA; Mr. Friedman, Naval Research Laboratory; Mr. Urey, University of California at La Jolla, California; and Mr. Fritz, U. S. Weather Bureau.

I think it is clear from the three sections of this symposium that there is no lack of problems for research in space. There are, however, a few problems that have not been mentioned in the course of yesterday and today.

One topic is that of γ-ray astronomy. We have talked about radio waves, X rays, ultraviolet radiation, and in principle one could also measure γ rays from colliding galaxies.

Does somebody want to discuss what can be done in this field?

Mr. Friedman: There is a definite program of γ-ray astronomy planned by Dr. Rossi and his colleagues at MIT, with the idea of looking for γ rays in the 100-Mev range from such sources as the Crab nebula.

We have been looking for harder radiation from the sun. I don't know that it qualifies as γ-ray astronomy, but these would probably be non-thermal radiations.

We find that in the daytime we can see a flux of about 10^{-7} erg/cm^2/sec between 20,000 volts and 50,000 volts. This radiation does not necessarily come from the sun, and it is in the range where it could conceivably be bremsstrahlung from the Van Allen radiation belt.

If we assume that we have about 5×10^5 oxygen atoms/cm^3 at a height of 1000 km, and assume a scale height of 500 km, then the expected bremsstrahlung flux, if we take the highest values quoted by Van Allen, is of the order of magnitude of our measured daytime flux. The obvious thing would be to try the same experiment at night. If the radiation disappeared at night then we could conclude that we have been looking at radiation from the sun.

Mr. Goldberg: I think Winckler has observed very intense short bursts of γ radiation from a balloon.

Mr. Friedman: There has been one observation by Winckler in connection with a class 2 flare, in which he detected a very sharp burst of γ rays in the range of a half million volts.

Mr. Spitzer: Meteorology is another area that has not been extensively discussed. We have a meteorologist on the panel this afternoon. I wonder if Dr. Fritz would like to say something about meteorological research in satellites.

Mr. Fritz: Actually several topics of meteorologic interest were mentioned today, such as the meteorology of Mars, which may give some clues about our own atmosphere. An observing station on the moon would yield information of great value. One observation from the moon that seems to be of great basic interest is that of the entire emission spectrum from the earth.

Mr. Jastrow: I should like to ask Mr. Fritz about the meteorological effects of a warm and dense zone in the auroral region. Between 150 and 180 km the temperature is several thousand degrees higher than at lower latitudes, and the air is denser. I should like to know what comments you have on the circulations produced by this distribution. Is it of any interest to you as a meteorologist?

Mr. Fritz: Definitely yes, because we really do not know all the interactions between the upper and lower atmospheres. There has been a great deal of speculation on the influence of abnormal solar emissions on events in the atmosphere.

Mr. Newell: I might comment here that between 40 and 50 km LaGow has shown that in the northern regions the atmosphere is warmer and of higher pressure, but there is very little change in density, relative to White Sands.

Mr. Friedman: I should like to ask Dr. Shapley what information we have now on the effects of solar flares on the upper ionosphere, meaning E region and F region. In the early part of the IGY program we conducted flare experiments with the idea of checking the hard X-ray radiation to see if it was there, knowing that we needed something in that range to produce the absorbing region.

As an aftermath of the eclipse experiment, we fired the last rocket as a background shot and by pure chance got it off in the middle of a class 2+ flare. This rocket was instrumented for soft X-ray measurements in the Lyman-α region. Although I would not want to quote the magnitude of the X-ray flux, it is definitely 2 or more times the normal X-ray flux that we would see in E region in the absence of a flare, and possibly 5 times as high.

Mr. Shapley: It is difficult to get information on the E and F region during solar flares by radio techniques, because the increased ionization in the D region masks the upper layers. There have been two instances I know of in which the electron density in the F region was shown to increase at the very beginning of the radio fade-out, before the fade-out became intensive enough to prevent echoes getting back from the F region. These were from two highly exceptional flares. I do not know whether this happens during an ordinary flare.

As far as the E region is concerned, all the evidence that there is, I think, is for no change.

Mr. Spitzer: I should like to raise a somewhat different question, directed primarily, I suppose, to Mr. Newell and Mr. Townsend, or to other members of the panel, and that is this: What basic limiting factor is there on the size of a telescope out in space? It is clear that on the earth's surface there is a limiting factor set by the deformation of a mirror under the force of gravity, plus the basic limitation of seeing, which limits what can be obtained from a big mirror.

What are the basic limitations in space? How big a telescope mirror can one conceive of?

Mr. Townsend: The answer depends on th time you allow for assembly in space, and fo the figuring of a large mirror.

Mr. Goldberg: If the telescope is very large you would probably have to do the final figuring of the surface in the satellite, because i you figure the mirror on the earth, under th influence of gravity, it will not have the sam shape in a gravity-free environment.

Mr. Newell: It would seem to me that th limitations at the present time are set by th engineering involved in getting the material into orbit and assembling them there.

I can throw the question partially right bac at you.

Isn't it true that the immediate need is t obtain precision in those regions in which cannot be obtained on the ground because c atmospheric distortions? For that purpose it not necessary to require much additional ligh collecting power.

Mr. Urey: I should like to ask the pan whether it is more feasible to make large telescopes on satellites about the earth, or to pr duce them on the moon. If they are to be a sembled on a satellite in space, by a very gre effort, then perhaps we will have to wait could, rather, wait until we have transport the moon.

Mr. Goldberg: The moon is not entirely fr of gravity. To the extent that gravity is limiting factor in deforming the surface of lar mirrors, then it would be preferable to use space station.

Mr. Urey: On the other hand, it might much more comfortable for the workmen have at least one-sixth of the earth's gravi around.

Mr. Menzel: The main advantage, it seen to me, that you get from a station in space, y also get from the moon, which is the freedc from the effects of the atmosphere. It would a considerable advantage to observe from so ground, even with a finite gravitational for

Mr. Newell: As a general comment to t question, I think that the engineering proble involved in either approach right now are formidable that we do not really apprecia them yet, and that we really cannot evalua which one is going to be the less difficult.

Low-Energy Cosmic-Ray Events Associated with Solar Flares

George C. Reid and Harold Leinbach

Geophysical Institute
University of Alaska
College, Alaska

Abstract—As a result of the IGY riometer program, it has been found that the measurement of ionospheric absorption in arctic regions is a sensitive method of detecting low-energy cosmic rays associated with solar flares. The normal morphology of these events is described, and details are given of the 24 such events that have been detected in the period from May 1957 through July 1959. Two features have been noted: an apparent asymmetry in the distribution of cosmic-ray-producing flares across the solar disk; a pronounced degree of uniformity in the distribution of the radio-wave absorption over the terrestrial polar cap. These features are discussed, and tentative explanations are suggested.

As part of the IGY program of ionospheric vestigation, the measurement of radio-wave sorption has been carried out at several high- titude stations using the riometer [*Little and einbach*, 1959], an instrument that continu- sly monitors cosmic noise at a preselected equency. The most important result of this ogram to date has been the discovery that e emission of fast particles from the sun ter a flare is a much more frequent occurrence an had hitherto been suspected. After the eat cosmic-ray-producing flare of February , 1956, very strong radio-wave absorption curred in high latitudes for about 3 days; it as noted by *Little and Leinbach* [1958] and scussed in some detail by *Bailey* [1957], who served the absorption on VHF ionospheric rward scatter links in the arctic. *Reid and llins* [1959] later gave details of two more these events, in July and October 1957, and a discussion of the various classes of abnor- al absorption events found in high latitudes olar blackouts) they proposed the name 'type I absorption' for them. They suggested that ost of the features of type III events could explained by assuming that the upper at- osphere was ionized by fast protons emitted om the sun after a major flare; this sugges- n has also been made independently by *iley* [1959] and *Hultqvist* [in press (*a*)].

The riometer can be shown to be an ex- emely efficient detector of vertically incident otons in the energy range from 5 to 50 Mev; rough calculations show that a flux of about 10 protons/cm^2/sec can readily be detected if the energy is about 10 Mev. It can be shown that, for a given amount of ionization, the radio-wave absorption is a maximum if the local ionospheric collisional frequency is equal to the angular frequency of the radio waves. For the riometer, operating at a frequency of 27.6 Mc/s, this occurs at a height of about 50 km; the ionization produced by protons whose penetration depth is either more or less than this value will cause less absorption than that due to protons reaching exactly this level.

Confirmation of the proton theory came from balloon observations of cosmic rays over Fort Churchill by Anderson [1958] during a type III event on August 22, 1958. These observations definitely showed the presence of a flux of protons at balloon heights whose energies were not high enough to allow penetration to ground level. Many such balloon flights have been made during subsequent events at both high and middle latitudes, notably by the University of Minnesota and State University of Iowa groups, and reference is made to them in Table 1. In the period from May 1957 through July 1959, 24 type III events have been observed by means of the riometer technique; this figure may be contrasted with the 5 occasions in the history of cosmic-ray recording on which increases of intensity associated with solar flares have been recorded at ground level. The relevant details of these type III events have been

TABLE 1—*Type III polar cap absorption events: May 1957–July 1959*

			Maximum absorption		Probable flare				References*	
Date	Starts, UT	Duration, greater than: hours	at 27.6 Mc/s, db	Solar activity at 27.6 Mc/s	Time, UT	Imp.	Helio-graphic position	Plage region	Radio data	Balloo satelli data
1957										
May 19	0200	10 (B)	1	N.S.† 0020–0045						
		6 (C)	1							
June 22	by 1000	x	x	. . .						
July 3	by 1100	46 (C)	6	. . .	0800	3+	N13 W40	4039	*a*	
July 24	2015	11 (C)	2	N.S. 1815–1935	1816	3	S24 W22	4070		
Aug. 29	1300	77 (C)	9	. . .	1031	3	S24 E22	4125		
Sep. 2	by 2100	32 (C)	9	N.S. 1440–2230	1313	3	S25 W36	4125		
Sep. 12	by 1200	18 (80°N,S)[b]	. . .	.	0709‡	2	N12 W15	4134	*b*	
			0.5 (C)							
Sep. 21	by 1930	31 (C)	5		1332	3	N13 W08	4152		
Sep. 26	by 2315	29 (C)	2	N.S. 1930–2315	1907	3	N26 E15	4159		
Oct. 21	by 0700	13 (C)	5		1637 (20th)	3+	S25 W45	4189	*a*	
1958										
Feb. 10	by 0700	30 (80°N,S)[b]	. . .		2108 (9th)	2+	S13 W14	4400	*b*	
		. . .	>12 (C)							
Mar. 25	by 2230	96 (FY)	12 (C)		0950 (23rd)	3+	S15 E80	4476	*c*	*c*
Apr. 10	1130	30 (C)	3 (C)		1010	1+	N18 W78	4485		
		40 (FY)	4.5 (FY)							
July 7	0130(T)	78 (C)	>15	SCNA§ 0030	0039	3+	N24 W02	4634	*d,e,f,g*	*h*
		120 (B)	>15	N.S. 0040–0210						
		x (T)	>15							
July 29	0405(T)	4 (C)	0.7	SCNA 0303	0303	3	S14 W43	4659		
		30 (B)	1	N.S. 0335–0450						
		22 (T)	1.5							
Aug. 16	0600(T)	x (C)	x	SCNA 0435	0432	3+	S14 W53	4686	*d,e,f*	*j*
		60 (B)	13	N.S. 0440–0515						
		56 (T)	>15							
Aug. 21	1500(B)	x (C)	3							
		19 (B)	3							
		x (T)	x							
Aug. 22	by 1700‖	80 (C)¶	>10	N.S. 1500–2100	1417	3	N21 W08	4708	*e,k*	*j,k,l*
		80 (B)¶	>9							
		80 (T)¶	9							
Aug. 26	0100(T)	57 (C)	>10	N.S. 0020–0145	0005	3	N20 W54	4708	*d,e,f*	*j*
		89 (B)	12							
		71 (T)	>13							
Sep. 22	1430(B)	x (C)	4		0741 ‡	2+	S17 W42	4765		
		68 (B)	4		1014	2	N18 W65	4756		
		82 (T)	4							
1959										
May 11	0130	92 (C)	>15	SCNA 2100(10th)	2055 (10th)	3+	N23 E47	5148	*m*	*m*
		200 (B)	>15	N.S. 2118–2315 (10th)						
		190 (T)	>15							
				SCNA 2010(11th)	2006 (11th)	2+	N08 E39	5148		
				N.S. 2020–2135 (11th)						
July 10**	0700(T)	90 (C)¶	>15	N.S. 2046–0400	1937(9th)	2+	N19 E67			
				SCNA 0215	0210 (10th)	3+	N22 E70			
July 14**	by 0700	51 (C)	>15	SCNA 0332 N.S. 0340–0700	0342	3+	N16 E07			
July 16**	by 2250	34 (C)	>15	SCNA 2119 N.S. 2125 0015	2115	3+	N08 W26			

* See lettered footnotes: *a*, *b*, etc.

† N. S. = solar noise storm.

‡ Flare identification dubious.

§ SCNA = sudden cosmic noise absorption.

‖ *Anderson* [1958] reports detection of protons at balloon heights as early as 1525 UT.

¶ Event still in progress at start of subsequent flare.

** *Brown* (private communication, 1959) has reported detection of particles at balloon heights over College throughout these events. Winckler (private communication, 1959) has reported similar observations over Churchill and Minneapolis.

(B) = Barrow, Alaska.

(C) = College, Alaska.

(FY) = Fort Yukon, Alaska.

(T) = Thule, Greenland.

a: Reid and Collins, 1959.

b: Hakura and others, 1958.

c: Freier and others, 1959.

d: Hultqvist, in press (*b*).

e: Hultqvist and Ortner, 1959.

f: Hultqvist and others, 1959.

g: Harang and Tröim, 1959.

h: Brown, 1959.

i: Leinbach and Reid, 1959.

j: Rothwell and McIlwain, 1959.

k: Anderson and others, 1959.

l: Anderson, 1958.

m: Ney and others, in press.

npiled in Table 1. The normal sequence of ents is as follows:

1. A major solar flare occurs, usually accomnied by a short-wave fadeout over the sun- hemisphere of the earth; these fadeouts are used by electromagnetic radiation from the re, and are recorded by the riometer if it ppens to be on the sunlit hemisphere. They referred to in the table as SCNA's (sudden smic noise absorption).

2. The flare is almost without exception folved by a strong low-frequency solar noise rm, often lasting for several hours; this is o recorded by the riometer when the sun is ove the horizon. The column in the table aded 'Solar activity at 27.6 Mc/s' refers ely to riometer observations made in Alaska, d the absence of data for some events does t mean that a noise storm did not occur.

3. Within a few hours after the flare, the type absorption sets in over the entire polar cap, actual onset often being obscured by the ar noise storm; the absorption reaches a maxum within a few hours, then decays slowly ring the following few days.

4. After a day or two a sudden-commencent magnetic storm almost invariably occurs, en accompanied by intense auroral activity.

The starting times listed in Table 1 should be ken as upper limits, since it is quite possible at more sensitive equipment would have own weak absorption occurring before these es. Similarly, the durations should be taken merely lower limits, since the decay of the sorption is quasi-exponential and a precise ration is undefinable.

The flare data have been taken from the PL tabulations of Solar-Geophysical Data ued monthly by the National Bureau of ndards.

Two important points can be noted from this lection of material:

1. Of the 24 events listed, the identification the corresponding flare is considered to be tain in 13 and highly probable in a further Examination of the heliographic positions of se 18 flares shows that 12 occurred to the st of the solar central meridian and 6 to the t. Assuming the entire solar-flare population during the period to be symmetrically distributed about the central meridian, the probability of this asymmetric distribution of cosmic-ray flares occurring by chance is only 7 per cent. Thus, although the statistics are not as yet really adequate, there is a clear indication that solar cosmic rays originating on the western half of the visible disk can reach the earth more readily than those from the eastern half.

This can be taken as favoring the hypothesis of the outward extension of a solar magnetic field by low-velocity charged particles traveling radially outward from the sun, as has been suggested by *Parker* [1958*a*]. The rotation of the sun will produce a curvature of the particle streams [*Chapman*, 1929] and consequently of the lines of force, which will be convex toward the west. The fast low-density protons suddenly ejected into this field will tend to travel along the lines of force, so that protons originating to the west of the solar central meridan can reach the earth more easily than those from the east. With the limited amount of data now available, it is impossible to do more than present this as a tentative suggestion.

2. The duration and intensity of these events at College are less than at the more northerly stations, and examination of the recordings from Farewell, to the south of College, shows that type III events, with a few exceptions, were either very weak or absent. Since the difference in geomagnetic latitude between College and Farewell is only 3°.3, this suggests that College is normally close to the southerly border of the region affected by protons penetrating only to the lower ionosphere, and that this border is very sharply defined. The few occasions on which strong type III absorption has been observed as far as 7° south (geomagnetically) of College occurred only after the onset of unusually strong magnetic storms of the sudden-commencement type. This behavior seems to indicate that on these occasions the geomagnetic field became sufficiently perturbed to allow the lower-energy protons to reach latitudes normally forbidden by a cutoff mechanism of the Störmer type [*Freier and others*, 1959].

However, there is a striking similarity in the form of the records from Barrow and Thule, which are separated in geomagnetic latitude by about 19°. This agrees with the observations of

Bailey [1959] on the February 23, 1956, event, and can be explained by a sharp flattening in the particle energy spectrum below the energy corresponding to a Störmer cutoff at 70° geomagnetic latitude. Another possible explanation, however, may lie in the confinement of the geomagnetic field within a finite region due to the impact of the streams of low-energy particles responsible for setting up the interplanetary magnetic field. Geomagnetic field lines from the polar regions, which would normally cross the equatorial plane in regions of low magnetic field intensity, would become very greatly disturbed under these conditions, and may link up with the solar stream field lines to preserve continuity of the field, thus providing a natural path for the fast protons. Since these geomagnetic field lines are the ones that intersect the earth to the north of the auroral zone [*Chapman and Ferraro,* 1931, 1932; *Parker,* 1958*b*], we might expect to see effects approximating those predicted by the Störmer theory of charged particles in a dipole field only to the south of the auroral zone. To the north we would expect to see more intense effects which would also be much more uniform, since the energy selection provided by the Störmer cutoff would no longer apply.

This suggestion is rather difficult to justify completely, but it does have the considerable merit of placing the low-latitude limit of the uniform polar cap absorption at the position of the auroral zone. Any theory that explains the uniformity by a flattening in the proton energy spectrum below a certain energy must also explain the remarkable coincidence that this energy nearly always corresponds to a Störmer cutoff at just the latitude of the auroral zone. Since the energy spectrum of the protons emitted from the sun can in no way be influenced by the geomagnetic field, such an explanation would be extremely difficult.

The work described here has been supported by the National Science Foundation under IGY contracts 6.20, 1.43, and Y.22.6. The initiation of the US IGY riometer program is due to Dr. C. G. Little, and the successful operation of the program would not have been possible without the constant support of the many field station operators.

References

Anderson, K. A., Ionizing radiation associat with solar radio noise storm, *Phys. Rev. Lette 1,* 335–337, 1958.

Anderson, K. A., R. Arnoldy, R. Hoffm. L. Peterson, and J. R. Winckler, Observati of low energy solar cosmic rays from the fl of 22 August, 1958, *State University of Io and University of Minnesota Report,* 1959

Bailey, D. K., Disturbances in the lower io sphere observed at VHF following the sc flare of 23 February 1956 with particular ref ence to auroral-zone absorption, *J. Geophys. I search, 62,* 431–463, 1957.

Bailey, D. K., Abnormal ionization in the lo ionosphere associated with cosmic-ray flux hancements, *Proc. IRE, 47,* 255–266, 1959.

Brown, R. R., Excess radiation at the Pfot maximum during geophysical disturbances, *Geophys. Research, 64,* 323–329, 1959.

Chapman, S., Solar streams of corpuscles: th geometry, absorption of light, and penetrati *Monthly Notices Roy. Astron. Soc., 89,* 4 470, 1929.

Chapman, S., and V. C. A. Ferraro, A new the of magnetic storms, Part I, the initial pha *Terrestrial Magnetism, 36,* 77–97, 171–186, 19 *37,* 147–156, 421–429, 1932.

Freier, P. S., E. P. Ney, and J. R. Winckl Balloon observations of solar cosmic rays March 26, 1958, *J. Geophys. Research, 64,* 6 688, 1959.

Hakura, Y., Y. Takenoshita, and T. Otsu Polar blackouts associated with severe geom netic storms on Sept. 13, 1957, and Feb. 1958, *Rept. Ionosphere Research Japan, 12,* 4 468, 1958.

Harang, L., and J. Tröim, An example of hea absorption in the VHF-band in the arctic io sphere, *Planetary and Space Science, 1,* 1 104, 1959.

Hultqvist, B., On the interpretation of ioni tion in the lower ionosphere occurring on b day and night side of the earth within a hours after some solar flares, *Tellus,* in p (*a*).

Hultqvist, B., On lower ionosphere electron tachment and recombination coefficients tained from measurements of nondeviative io spheric absorption, *Arkiv Geofysik,* in press (

Hultqvist, B., and J. Ortner, Observations of tense ionization of long duration below 50 altitude after some strong solar flares, *Nat 183,* 1179–1180, 1959.

Hultqvist, B., J. Aarons, and J. Ortner, Rep on and interpretation of strong ionization the lower ionosphere, occurring in high latitu within a few hours after some solar flares, *Rept. 1, Contract AF 61(514)–1314, Kiruna G physical Observatory,* February 1959.

Leinbach, H., and G. C. Reid, Ionization of upper atmosphere by low-energy charged particles from a solar flare, *Phys. Rev. Letters, 2,* 61–62, 1959.

ttle, C. G., and H. Leinbach, Some measurements of high-latitude ionospheric absorption using extraterrestrial radio waves, *Proc. IRE, 46,* 334–348, 1958.

ttle, C. G., and H. Leinbach, The riometer—a device for the continuous measurement of ionospheric absorption, *Proc. IRE, 47,* 315–320, 1959.

ey, E. P., J. R. Winckler, and P. S. Freier, Protons from the sun on May 12, 1959, in press.

rker, E. N., Dynamics of the interplanetary gas and magnetic fields, *Astrophys. J., 128,* 664–676, 1958*a*.

Parker, E. N., Interaction of the solar wind with the geomagnetic field, *Phys. Fluids, 1,* 171–187, 1958*b*.

Reid, G. C., and C. Collins, Observations of abnormal VHF radio wave absorption at medium and high latitudes, *J. Atmospheric and Terrest. Phys., 14,* 63–81, 1959.

Rothwell, P., and C. McIlwain, Satellite observations of solar cosmic rays, *State University of Iowa Report,* 1959.

(Manuscript received September 8, 1959.)

Cosmic-Ray Measurements in the Vicinity of Planets and Some Applications[1]

Part I: Primary Cosmic Radiation

S. F. Singer and R. C. Wentworth

University of Maryland
College Park, Maryland

Abstract—The variation of the primary cosmic-ray intensity is calculated as a function of distance from a dipole in its equatorial plane. Different values of spectrum exponent and of minimum momentum are chosen. Results are presented in a general form but are also specialized to the case of the moon.

Scientific applications are indicated, such as determination of the cosmic-ray energy spectrum and low-energy cutoff, cause of the 'knee' and of Forbush decreases, value of the free-space cosmic-ray flux, and determination of magnetic fields of the moon and the planets.

The present paper deals only with the primary cosmic rays; the cosmic-ray albedo which produces the 'radiation belt' is discussed in Part II.

Introduction

Cosmic-ray measurements in the vicinity of lanets have many important scientific applicaions; for example, they can establish very diectly the existence or nonexistence of a magetic field of a planet. In the case of the earth e know the magnetic field and its dipole monent, and measurements in the immediate icinity of the surface and as a function of titude have been used to establish the energy pectrum of the primary cosmic rays (more preisely, they have been used to establish the igidity spectra of the various components of he primary cosmic rays—the protons, helium uclei, and heavier primary nuclei) [*Singer* 958a]. In spite of many measurements in balons and rockets, and now in earth satellites, ncertainties still exist concerning the exact rm of the spectra, particularly at the very w energy end but also beyond a particle gidity of about 60 Bv (rigidity is defined as omentum divided by electric charge) where e analyzing effect of the earth's magnetic field ases to be of value. Reasons for the uncerinties are partly instrumental, but they are chiefly due to the influence of cosmic-ray albedo (secondaries produced in the atmosphere and directed upwards) on measurements of the 'primary' intensity. At high latitudes there are uncertainties due to the complicating effects of the earth's quadrupole field and to our inadequate knowledge of the shadow cone at high latitudes; there is also some doubt as to whether the earth's magnetic field is really a dipole field to a sufficient approximation very far out from the earth.

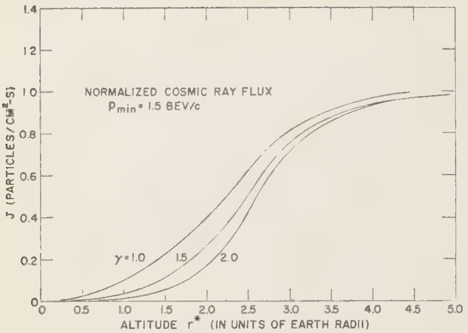
Because of these disturbing effects an alternate method for measuring the primary cosmic-ray spectra has been suggested by *Singer* [1955] which would consist of measuring the intensity increase with altitude, especially above the earth's equator where the variation can be calculated exactly from a theoretical point of view.

The purpose of this paper is mainly to calculate this increase in primary flux in the equatorial plane with distance from the center of the dipole for a range of rigidity spectra and for various values of the low-energy cutoff. These results can serve for comparison with any experimental determination of the primary cosmic-ray variation with altitude.

Scientific Applications

The applications of such measurements are numerous, and we shall discuss them in order.

[1] Physics Department Technical Report No. 114. pported in part by the U. S. Army Ballistics issiles Agency under Contract DA-36-034-ORD-31.



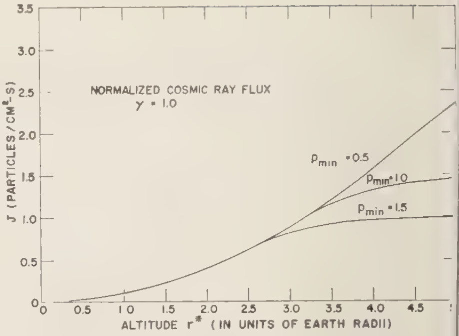


Fig. 1—J, the omnidirectional primary flux entering a sphere of unit cross section per second, is plotted as a function of a dimensionless variable $r^* = (r/R_E)(M_E/M)^{1/2}$ where R_E is 6.37 × 10^8 cm and M_E = 8.1 × 10^{25} gauss cm³. The curves are calculated for three different exponents γ of the spectrum and are normalized to give a flux of 1 at infinity. The curves have been drawn for a transparent dipole. The actual shadow cone of a planet can be taken into account as shown in Fig. 5.

Fig. 2—The variation to be expected in J f $\gamma = 1.0$ and three values of the low-energy cu off p_{min} = 0.5, 1.0, and 1.5 Bev/c are indicate The curves are normalized so that the p_{min} = 1 Bev/c curve gives a flux of 1 at infinity.

Determination of the spectrum of primary cosmic rays—The variation of cosmic-ray intensity with altitude or distance from the earth should serve as an excellent means of determining the momentum spectrum. The analysis of equatorial data is complicated by albedo. If the primary spectrum follows a power law with exponent 1.5, as suggested by *McDonald* [1958] and *Webber* [1958], then the experimental curve will deviate considerably from the 1.1 value assumed previously. Figure 1 shows the range of variation depending on the spectrum exponent γ. Albedo has some effect even at altitudes of several earth radii because of the trapping effect of the earth's magnetic field [*Singer*, 1958b; *Van Allen and others*, 1959]. However, we can discriminate between primaries and trapped albedo in three ways: (1) by using a Cerenkov detector, (2) by measuring only particles with charge $z > 1$, and (3) by measuring the (horizontal) directional flux *along* the line of force.

Determination of the intensity of cosmic rays in interplanetary space—Such measurements are of space-medical as well as astrophysical interest. As will be discussed, there is considerable difference between various theories of cosmic-ray time variations as to what this intensity is.

Determination of the low-energy cutoff of the cosmic-ray radiation—The measurements would establish the existence (Fig. 2) and, in particula the position, of the low-energy cutoff. Prese evidence indicates that a low-energy cutoff the cosmic-ray spectrum exists during periods high solar activity [*Neher*, 1956]. Howeve because of technical difficulties with the measu ments, we have no detailed knowledge of t day-to-day variations in the cutoff. In order establish the cause of the low-energy cutoff is certainly desirable to obtain a closer correlati between cosmic-ray measurements and t activity of the sun.

Investigation of the cause of the low-ener cutoff—One clear-cut answer which such measu ments can give concerns the possibility that t low-energy cutoff is an entirely local effe According to *Parker* [1956] it could be produ by a cloud of solar gas held close to the ea by its gravitational field. If the cloud w highly turbulent, low-energy cosmic rays wo have long diffusion transit times and would effectively screened from the earth. This mo of the knee and other models are discus elsewhere [*Singer*, 1958a]. However, Figure shows how the cosmic-ray intensity wo behave with altitude if such a screening cl were indeed located in the immediate vicinity the earth (about 2 to 3 earth radii out).

Investigation of the cause of cosmic-ray decre associated with magnetic storms—*Parker* [19 has also suggested that Forbush decreases due to the presence in the vicinity of the ea of a highly turbulent cloud which produ

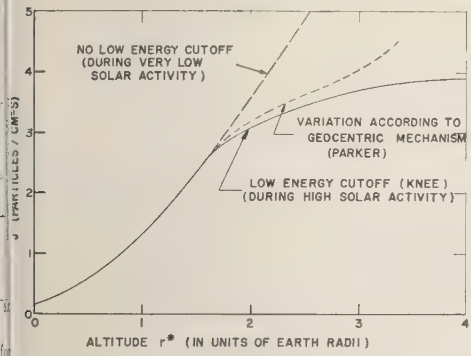


Fig. 3—The increase in total flux as a function f altitude at the geomagnetic equator. The bscissa is given in terms of earth radii. The solid ne shows the normally expected variation calcu- ated from the primary spectrum and geomagnetic heory. The short dashes show the variation to be xpected according to the geocentric mechanism f *Parker* [1956]. The long dashes show the varia- on to be expected in the absence of a low-energy utoff.

irly effective magnetic screening. Under this odel the altitude dependence of flux should be shown in Figure 4, whereas under other models e cosmic-ray intensity would be depressed in alue not only at sea level but also throughout terplanetary space. Thus the proposed measure- ents provide a clear-cut decision between ffering theories.

Examination of the falloff of the earth's magnetic ld—It has been suggested that the earth's agnetic field is screened by the motion of the terplanetary gas or other dynamic effects. The reening limit has been set as low as 5 earth dii and as high as 20 earth radii. Cosmic-ray easurements with altitude could provide a ear-cut answer here. In case of a cutoff at earth radii, for example, there should be no ther variation of intensity with increasing stance. In addition, more frequent measure- nts of cosmic-ray variation with altitude uld establish changes in the behavior of the th's magnetic field at these altitudes. Such asurements should be correlated, of course, th direct measurements of the magnetic d made with instruments sensitive enough measure such small field values.

nvestigation of the possible existence of magnetic ds for the moon and planets—This is one of the important applications of the method described in the present paper. (The planetary radiation belt of trapped albedo provides a striking demonstration of the existence of a magnetic field but does not measure its value.) As the cosmic-ray detector approaches the vicinity of the moon or the planet, the primary intensity falls off for two reasons. It will decrease because of the shadowing of a portion of the solid angle by the solid mass of the planet, and it may also decrease because of the existence of a magnetic field. It develops that this may be quite a sensitive detector for the existence of a dipole field. The results shown in Figure 1 and equation (15) can be used for this purpose. As a specific example, we show in Figure 5 the cosmic-ray variation observed when one approaches the moon; the spectrum used here has an exponent of 1.1.

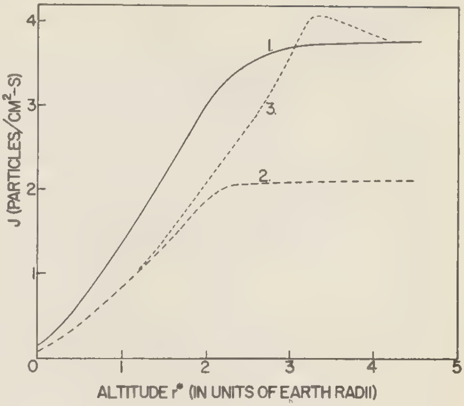


Fig. 4—The increase in total flux as a function of altitude at the geomagnetic equator. The abscissa is given in terms of earth radii. The solid line (1) shows the normally expected variation calculated from the primary spectrum and geomagnetic theory. The dashed line (2) shows the flux variation during a Forbush decrease; the cosmic-ray intensity is supposed to be depressed at large distances from the earth as well. The dotted curve (3) shows the flux variation expected from the geocentric theory of *Parker* [1956]; it may be possible even to observe an additional increase due to cosmic rays which are reflected by the geocentric barrier.

Some typical results are given. If the moon has a surface field of 1.1 gauss (corresponding to a magnetic moment of 5.8×10^{24} gauss cm^3), then the cosmic-ray intensity would drop off

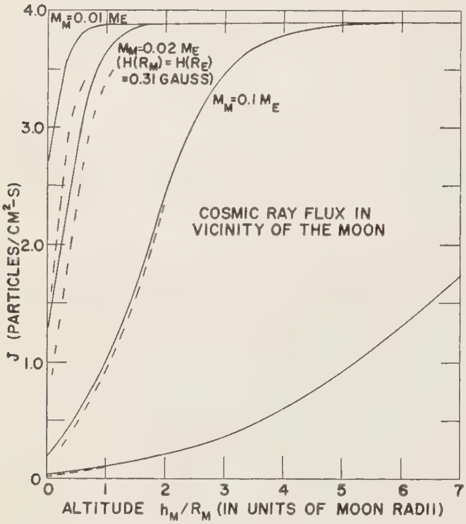


FIG. 5—Specific applications to the case of the moon. The expected omnidirectional primary flux is shown entering a sphere of unit cross section per second as it approaches the assumed lunar dipole in the equatorial plane. Various dipole moments are taken to show the variations. When $M_M = (1/50)M_E$ the surface value of the field at the lunar equator is 0.31 gauss—about equal to that of the earth. The solid curves are computed for a transparent moon, the dotted curves for a shadow cone which is equal to the geometric solid angle subtended by the moon. The bottom curve corresponds to $M_M = M_E$.

to the same value as that observed at the top of the earth's atmosphere (at 50 miles) at the equator. In other words, we would trace out completely the curve shown in Figure 1. If the moon had a surface field of 0.1 gauss, then we would observe the curve down to the low-energy cutoff point. In other words if there are no low-momentum ($<$ 1.5 Bev/c) particles in the interplanetary space we cannot measure the moon's field to a value less than 0.1 gauss. Assuming, however, that lower-energy particles do exist, and that our detectors are sensitive down to a particle energy of 50 kev (corresponding to a momentum for protons of 10 Mev/c), then we can measure a lunar surface field of slightly less than 1 milligauss, corresponding to a magnetic moment of 3.9 × 10^{21} gauss cm^3.

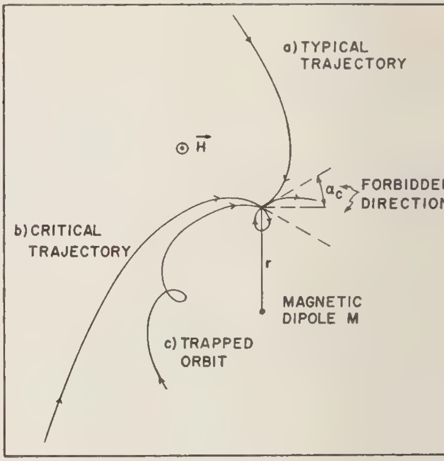


FIG. 6—Some typical orbits of cosmic rays the equatorial plane of a dipole. The existence forbidden directions of approach is illustrated the fact that only particles in orbit c can en with angle less than α_c, and such particles cann come in from infinity, being in trapped orbits

In the calculations of this paper it has been assumed that at infinity the cosmic rays isotropic and have an integral moment spectrum $N(> p) = Cp^{-\gamma}$/cm^2 s ster. W the cutoff momentum $p_{\min} = 1.5$ Bev/c, th values for γ have been used: $\gamma = 1.0, 1.5$, and 2 In this case the total intensity at infinity been normalized to 1. Also, for the case $\gamma = 1.0$, two additional values of $p_{\min}$ ha been taken: $p_{\min} = 0.5$ and 1.0 Bev/c.

Cosmic-Ray Flux in Dipole Field

Even with the simple field of a magne dipole it is not possible to obtain the solut for the equation of motion of a charged parti in closed form. However, it is possible to disc the problem in general terms. For one thing, field has the effect of turning particles of l energy away before they penetrate too close the dipole itself. Therefore, if it is assumed t at infinity the cosmic rays have a continu energy spectrum, the intensity of the cosm rays will decrease as the dipole is approach

This decrease in the intensity as the dipol approached is quite difficult to calculate in general case. However, an exact solution to t problem can be obtained for the special c where the dipole is approached along its eq rial plane. In this case, if the form of the smic-ray spectrum at infinity is assumed, it possible to obtain the exact dependence of e intensity on the distance from the dipole id on the dipole strength. Also, it is possible to lculate the dependence of the altitude disibution on the spectrum assumed for the smic rays at infinity.

The simple Störmer theory of the motion of charged particle in the field of a magnetic pole [for example, *Montgomery,* 1949] leads an expression for a critical angle α_c (Fig. 6). articles having a momentum p can come in om infinity to a distance r from the dipole ith angles α greater than α_c. The path of such articles is reversible in a sense. That is, if a article of opposite charge is started from r in e opposite direction with the same momentum, will retrace the path used in coming in from finity. It is easy to see why angles less than come to be forbidden to such particles. If a article is started with such an angle it can be own to be in a trapped orbit, looping indefitely about the dipole (Fig. 6). Since this article never reaches infinity, it must follow at no particle from infinity can get into such orbit if it has an angle at r less than α_c. This ementary treatment, of course, neglects any attering of particles near the critical orbit into ch trapped orbits.

The above discussion is concerned with partis moving entirely in the equatorial plane of e dipole. If the point at r is still in the equarial plane of the dipole but the charged rticles are not restricted to move in the uatorial plane, the critical angle α_c becomes e half-angle of a cone, called the forbidden ne [*Montgomery,* 1949]. Thus particles hav- a momentum p and coming in from infinity r can arrive at that point with only a limited nge of directions determined by the critical gle α_c.

According to Liouville's theorem, the inteny of particles having a given constant montum p is constant along its trajectory. Thus, the particles are isotropic at infinity, the io of the intensity at r to that at infinity will simply $1/(4\pi)$ times the solid angle of the owed cone.

Therefore, if the spectrum of the cosmic rays known, the ratio of the counting rates of omnidirectional counters close to the dipole and at infinity can be easily calculated with an equation which is derived in the Appendix.

Of course the counting rate at the surface of a planet will be one half of the calculated value, since the planet will intercept half of the particles before they reach the counter. As the counter moves away from the planet the effective solid angle subtended by the planet will decrease, and the counting rate will approach the calculated value.

Appendix

Let us assume that the integral spectrum of cosmic rays at infinity is

$$N(>P) = Cp^{-\gamma}/\text{cm}^2\text{ s ster} \qquad (1)$$

where p is the momentum of the protons in units of Bev/c. The cosmic rays at infinity are taken to be isotropic.

To calculate the forbidden cone in the equatorial plane of a dipole, we note [*Montgomery,* 1949] that

$$\alpha_c = \cos^{-1}[2/r_0 - 1/r_0^2]$$

where

$$r_0 \equiv r/C_{st} \qquad (2)$$

and

$$C_{st} \equiv [eM/p]^{\frac{1}{2}}$$

Here e is the charge of the electron, 1.6×10^{-20} emu; M is the dipole magnetic moment in gauss cm^3; p is the momentum of the proton in gm cm/s; and C_{st} is the Störmer unit of length in cm.

Now, in equation (2) we introduce the additional dimensionless quantities

$$\mu \equiv M/M_E; \qquad \rho \equiv r/R_E \qquad (3)$$

where M_E is the dipole moment of the earth, μ is the ratio of the dipole moment of the planet to that of the earth, and ρ is the ratio of the distance from the dipole to the radius of the earth. Thus we find

$$C_{st} = 1.139 \times 10^3\mu^{1/2}p^{-1/2} \qquad (4)$$

Solving (2) for C_{st} yields

$$C_{st} = r(1 + \sqrt{1 - \cos\alpha_c}) \qquad (5)$$

where the positive sign is taken because a decrease of α_c implies a larger momentum of the protons involved.

Substituting (4) in (5), we can solve for p.

$$p_{gc} = \frac{(1.139 \times 10^3)^2\mu}{C_{st}^{\ 2}}$$

$$= \frac{(1.139 \times 10^3)^2\mu}{r^2(1 + \sqrt{1 - \cos\alpha_c})^2}\left[\frac{\text{gm cm}}{\text{s}}\right] \qquad (6)$$

A subscript gc has been added to p since this value of p is the geomagnetic cutoff; that is, the minimum momentum a particle can have to come in from infinity and arrive at the angle α_c. Of course, a particle having momentum p_{gc} can arrive at any angle greater than α_c (within the allowed cone).

However, p_{gc} is in gm cm/s and we wish to express it in Bev/c. Now,

$$1 \text{ gm cm/s} = 1.873 \times 10^{13} \text{ Bev/c} \qquad (7)$$

Then, substitution of (3) and (7) into (6) yields

$$p_{gc} = \frac{59.6\mu}{\rho^2(1 + \sqrt{1 - \cos\alpha_c})^2}\ \text{Bev/c} \qquad (8)$$

It can be seen that if μ and ρ are both equal to 1 and $\cos\alpha_c$ is $+1$, then p_{gc} is 59.6 Bev/c, a well-known result of geomagnetic theory.

Now, from (1), the counting rate per second at infinity of an omnidirectional counter of unit cross-sectional area is given by

$$J_\infty = \int_0^\pi 2\pi\, d\alpha \sin\alpha N(>p_{\min}) = \frac{4\pi c}{p_{\min}^\gamma} \qquad (9)$$

To calculate the counting rate of a unit omnidirectional counter at some point at a distance r from the dipole in its equatorial plane, we must apply (8). Consider all particles which enter the counter at some angle between α and $\alpha + d\alpha$. The solid angle contained between α and $\alpha + d\alpha$ is $2\pi \sin\alpha\, d\alpha$. Since the particle of minimum momentum which can enter the counter at an angle α is given by (8), the number entering the counter per second in this solid angle is

$$dJ = 2\pi \sin\alpha\, d\alpha C p_{gc}^{\ -\gamma}, \text{ for } p_{gc} \geq p_{\min}$$

$$= 2\pi \sin\alpha\, d\alpha C (p_{\min})^{-\gamma}, \text{ for } p_{gc} \leq p_{\min} \quad (10)$$

Therefore, the counting rate of the counter is

$$J = \int_{\alpha=0}^{\alpha_m} \frac{2\pi \sin\alpha\, d\alpha C}{p_{gc}^\gamma} + \int_{\alpha=\alpha_m}^{\pi} \frac{2\pi \sin\alpha C}{p_{\min}^\gamma} \qquad (1$$

where α_m is determined from

$$p_{\min} = 59.6\mu\rho^{-2}(1 + \sqrt{1 - \cos\alpha_m})^{-2} \quad (1$$

Finally, J becomes

$$J = 2\pi C\left[\int_{\alpha=0}^{\alpha_m} \frac{\sin\alpha\, d\alpha}{p_{gc}^\gamma} + \frac{1 + \cos\alpha_m}{p_{\min}^\gamma}\right]$$

$$= 2\pi C\Bigg[(59.6\mu/\rho^2)^{-\gamma}$$

$$\cdot \int_{\alpha=0}^{\alpha_m} (1 + \sqrt{1 - \cos\alpha})^{2\gamma} \sin\alpha\, d\alpha$$

$$+ \frac{1 + \cos\alpha_m}{p_{\min}^\gamma}\Bigg] \qquad (1$$

Now, let $a = 1 - \cos\alpha$; $da = +\sin\alpha\, d\alpha$. $\alpha = 0$, $a = 0$; if $\alpha = \alpha_m$, $a_m = 1 - \cos$ Then

$$J = 2\pi C\Bigg[(59.6\mu/\rho^2)^{-\gamma} \int_{a=0}^{a_m} (1 + a^{\frac{1}{2}})^{2\gamma}\, da$$

$$+ \frac{1 + \cos\alpha_m}{p_{\min}^\gamma}\Bigg] \qquad (1$$

This is our final result, which can now be eval ated for different energy spectra.

References

McDonald, F. B., Study of primary cosmic-ray and proton energy spectra, geomagnetic cut energies and temporal variations, *Nuovo cimen Suppl., 8,* 500–507, 1958.

Montgomery, D. J. X., *Cosmic Ray Physi* Princeton University Press, 312–329, 1949.

Neher, H. V., Low-energy primary cosmic-r particles in 1954, *Phys. Rev., 103,* 228–236, 19

Parker, E. N., Modulation of primary cosmic-r intensity, *Phys. Rev., 103,* 1518–1533, 1956.

Singer, S. F., A new method for measuring t low-energy spectrum of primary cosmic ra *Phys. Rev., 98,* 1547, 1955.

Singer, S. F., The primary cosmic radiation and time variations, *Progress in Elementary Parti and Cosmic Ray Physics,* No. Holland Pu Co., Amsterdam, 203–335, 1958a.

Singer, S. F., 'Radiation Belt' and trapped cosm ray albedo, *Phys. Rev. Letters, 1,* 171, 1958b; also, Trapped albedo theory of the radiation belt, *Phys. Rev. Letters, 1,* 181, 1958b.

'an Allen, J. A., C. E. McIlwain, and G. H. Ludwig, Radiation observations with Satellite 1958ε, *J. Geophys. Research, 64,* 271–286, 1959.

Webber, W. R., The charge composition and energy spectra of primary cosmic-rays and the energy balance problem, *Nuovo cimento Suppl., 8,* 532–545, 1958.

(Manuscript received January 12, 1959.)

Aurora-Like Radar Echoes Observed from 17° Latitude[1]

R. B. Dyce, L. T. Dolphin, R. L. Leadabrand, and R. A. Long

Communication and Propagation Laboratory
Stanford Research Institute
Menlo Park, California

Abstract—Anomalous echoes are regularly observed by a shipborne radar located at Antigua, British West Indies. These echoes, observed at 32 and 140 Mc/s, have many of the characteristics of echoes from the auroras observed in the arctic, although visible auroras should not be observable at Antigua more frequently than once in 7 years. Similar observations at Stanford University indicate a correlation with one kind of sporadic-*E* ionization.

A shipborne, high-power radar (see Fig. 1) ocated at Antigua, British West Indies (17.2°N, 61.8°W, geomagnetic latitude 30°, magnetic dip 0°), regularly obtains anomalous echoes at anges from about 200 or 600 km on 32 Mc/s and occasionally on 140 Mc/s (see Table 1). These echoes resemble auroral echoes obtained in the arctic in the following features: (1) a endency for the echoes to occur in those directions where a radio ray is perpendicular to the earth's magnetic field; (2) a rapid, irregular fluctuation rate of the order of 10 or more; (3) sporadic occurrence in time and position; (4) a endency to occur at about 100 km height; (5) a wavelength dependence of equivalent cross ection of $\sigma_{32}/\sigma_{140} = \lambda^{5\pm3}$.

These echoes are unlike arctic auroral echoes or the following reasons:

1. Their occurrence (exceeding 10 per cent f the time and occurring each day) is far in xcess of extrapolation from the arctic based on robability of visual aurora. *Vestine* [1944] estimates the frequency of days with occurrence of auroras at 0.04 per cent, or 1 day in 7 years.

2. Poor correlation with concurrent geomagnetic disturbance was indicated by local earth-potential fluctuations and North Atlantic Quality announcements broadcast by WWV.

3. Visible auroras were not observed even when unusually strong night-time radar echoes were present.

A range vs. azimuth, intensity-modulated display is shown in Figure 2. This was an unusually strong event and consequently occurs in more directions than are typically observed. The U-shaped pattern suggests that the echoes lie either along a line of constant magnetic latitude or in those positions most closely orthogonal to the magnetic field lines.

By scanning the antenna in elevation, the echoes were shown to be occurring in the *E* region and not the *F* region.

Unlike arctic auroral echoes, there appears to be a lack of correlation with geomagnetic disturbance. Earth-potential disturbances [*Hessler and Wescott*, 1959] commenced immediately at Antigua with the Special World Interval that began at 0800, July 15, 1959. These indications of geomagnetic disturbances continued with only slightly reduced magnitude throughout the entire day. During this same period, unusually strong 100-Mc/s auroral echoes were being observed from Stanford (37.4°N, 122.2°W, geomagnetic latitude 44°, magnetic dip 62°). At Antigua, a search for aurora-like echoes was conducted during the evening of July 15. A few such echoes were found about 2155 local time but were not abnormally strong, and they disappeared within a half hour. On the other hand, the following day (July 16), strong aurora-like echoes were observed at noon and for several hours in the evening after the geomagnetic disturbance had largely subsided. These evening echoes were exceedingly strong on 32 Mc/s and were even observable on 140 Mc/s.

A night-time occurrence is suggested because

[1] This work was done under the support of Air Force Cambridge Research Center Contract Number AF 19 (604)-5209.

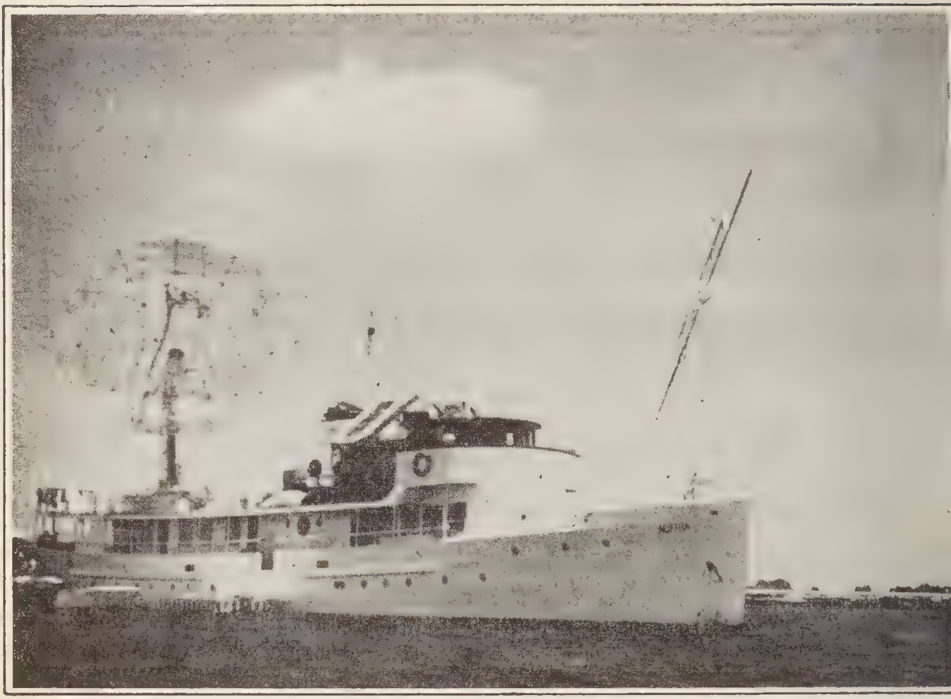


FIG. 1.

on many occasions the echoes were observed to begin in the early evening hours (20^h local) and to vanish about midnight. Echoes have also been seen near local noon, however.

These echoes resemble *E*-region echoes from magnetic-field-aligned ionization discovered at Stanford University [*Peterson and others,* 1955; *Leadabrand,* 1955; *Gallagher,* 1956], where this type of rapidly fluctuating echo was found to b associated with one type of sporadic *E*. Mor recently, field-aligned ionization at latitude similar to that of Stanford has been discovere by *Smyth and others* [1958] on an oblique 200 Mc/s scatter path.

The observed aspect-sensitive behavior indi cates that the echoes may be occurring fron

TABLE 1—*Characteristics of radars aboard motor vessel Acania anchored at Antigua, B. W. I.*

	32 Mc/s	140 Mc/s
Peak transmitter power	60 kw	60 kw
Antenna gain*	20 or 13 db (8-element Yagi)	250 or 24 db (30-ft-diameter dish)
Antenna beamwidth	45°	15°
Receiver bandwidth	6 kc	6 kc
Receiver sensitivity at antenna	2×10^{-16} watt	0.5×10^{-16} watt
Pulse width	300 μsec	300 μsec
Pulse repetition frequency	15 or 30 cps	15 or 30 cps

*These radars are capable of simultaneous operation with antennas synchronized and steerable in a directions.

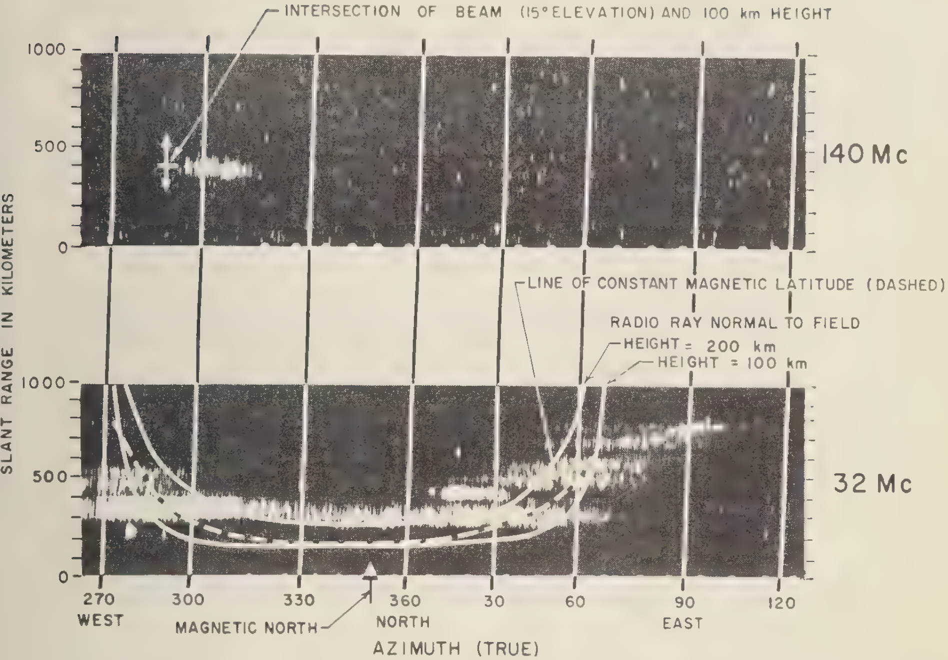


Fig. 2—Azimuth scan of aurora-like echoes obtained at Antigua, B. W. I., July 16, 1959.

scattering columns of ionization aligned with he earth's magnetic field. Such irregular ioniza-ion could well be caused by magnetic-field-guided particles in the same fashion as the for-nation of arctic auroras. However, long columns of field-aligned ionization are also a natural onsequence of magnetic control of charged articles initially distributed by isotropic turbu-ence. [*Booker, Gartlein, and Nichols,* 1955]. lthough the cause of these magnetic field-ligned ionized irregularities is not known, their xistence at such a low geomagnetic latitude is emarkable.

It is hoped that additional investigations at igher and lower frequencies (radars are avail-ble on the same ship which operate at 4.7, 11.6, 70, and 780 Mc/s will provide more wave-ength-dependence data to aid in an understand-g of the production mechanism of the ioniza-ion. The association with sporadic *E* also needs urther examination.

Riometers [*Little and Leinbach,* 1959] are currently being operated at Antigua on 30, 60, and 120 Mc/s. The 30-Mc/s riometer occasionally registers sporadic occurrence of up to 1 db of night-time absorption that may be related to the occurrence of these radar echoes. It is well known that ionospheric absorption occurs simultaneously with arctic aurora, most probably a result of *D*-region ionization produced by secondary X-rays.

References

Booker, H. G., C. W. Gartlein, and B. Nichols, Interpretations of radio reflections from the aurora, *J. Geophys. Research, 60,* no. 1, March 1955.

Gallagher, P. B., Analysis of a new type of radio scattering from the ionospheric *E*-region, *Radio Propagation Lab. Stanford Univ. Tech. Rept. 107,* May 7, 1956.

Hessler, V. P., and E. M. Wescott, Rapid fluctuations in earth-currents at College, *Geophys. Inst. Univ. Alaska Sci. Rept. 1,* ARCRC Contract no. AF 19(604)-3075, January 1959.

Leadabrand, R. L., Radio echoes from auroral ionization detected at relatively low geomagnetic latitudes, *Radio Propagation Lab. Stanford Univ. Tech. Rept. 98,* December 9, 1955.

Little, C. G., and H. Leinbach, The riometer—a device for the continuous measurement of ionospheric absorption, *Proc. IRE, 47,* 315–319, 1959.

Peterson, A. M., O. G. Villard, R. L. Leadabrand, and P. B. Gallagher, Regularly observable aspect-sensitive radio reflections from ionization aligned with the earth's magnetic field and located within the ionospheric layers in middle latitudes, *J. Geophys. Research, 60,* 492–512, 1955.

Smyth, J. B., J. L. Heritage, and S. Weisbrod, paper presented at meeting of URSI–IRE, Fall 1958.

Vestine, E. H., The geographic incidence of aurora and magnetic disturbance, northern hemisphere, *Terrestrial Magnetism and Atmospheric Elec., 49,* 77–102, 1944.

Studies of Magnetic Field Micropulsations with Periods of 5 to 30 Seconds[1]

W. H. CAMPBELL[2]

Institute of Geophysics
University of California
Los Angeles, California

Abstract—Magnetic field micropulsations with periods of 5 to 30 sec were studied for 7 months of 1958 at a station in southern California with a 2-m-diameter coil antenna of 21,586 turns. The local diurnal amplitude fluctuation attained maxima at 0945 and 1400 hours. Twenty-seven-day solar dependence and correlations with magnetic and ionospheric *F*-layer disturbances were evident. The storm time variation for micropulsation storms showed a secondary maximum at 65 min.

Introduction—Micropulsations of the earth's magnetic field have been a curiosity ever since *Stewart's* [1861] first observation of them a hundred years ago. More recently *Holmberg* [1953], *Angenheister* [1954], *Troyickaya* [1955], and *Kato and Wantanabe* [1957] have made significant studies of the pulsations. Oscillation periods in the broad range from several minutes to a number of seconds have been reported. There appear to be two groups, with periods of

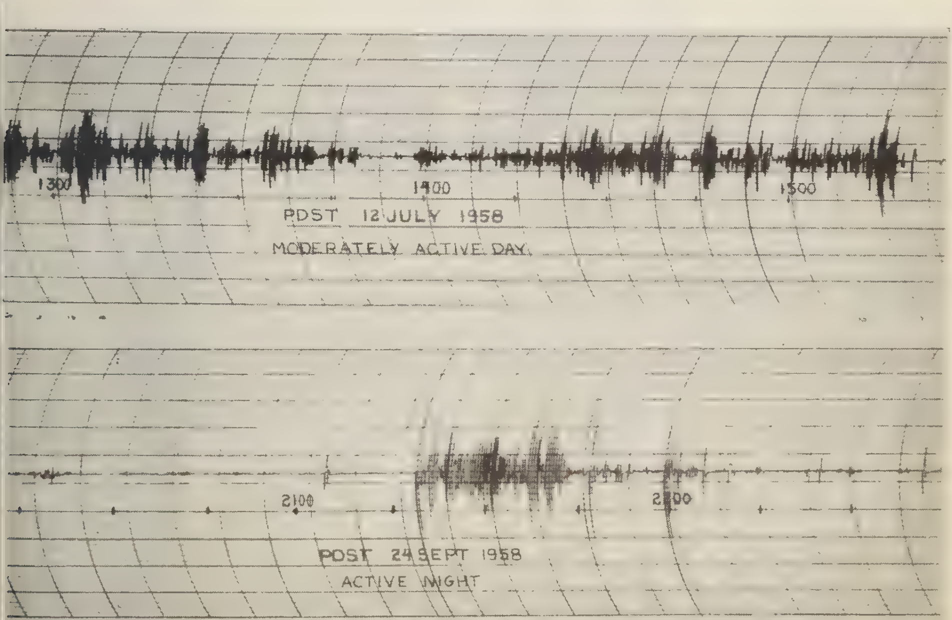


FIG. 1—Example of north-axis coil-antenna measurements.

[1] Assisted by the Office of Naval Research through Contract Nonr 233(47).

[2] At present with the Geophysical Institute, University of Alaska, College, Alaska.

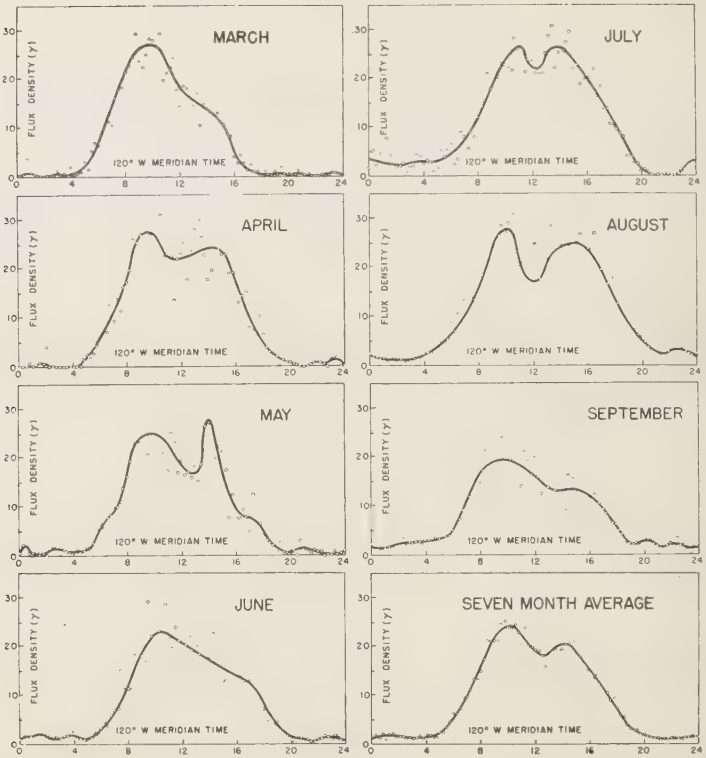


FIG. 2—Average diurnal behavior of micropulsations.

roughly 1 to 3 min and 5 to 30 sec, which have quite different average diurnal behaviors. The magnetic flux densities of the micropulsations are usually of the order of 1/3 gamma in magnitude. However, giant micropulsations with periods of 46 to 152 sec may reach 40 gammas in amplitude [*Sucksdorff,* 1939].

This report is concerned with the oscillations with periods of 5 to 30 sec which were measured with a north-axis coil antenna of 21,586 turns and 2 m in diameter at a Borrego, California, desert site (33°21.5′N, 116°17′W). The detection system had a band pass with 3 db points at 0.04 and 0.4 cycle/sec and a limiting sensitivity of 0.02 gamma. The magnetic field micropulsations, which were about ten-millionths of the earth's usual dipole field, generally appeared in groups of four to six oscillations reminiscent of a beat-type pattern. Maximum amplitudes were attained near local noon. The amplitude and duration of the oscillatory field were related to the magnetic planetary 3-hour range index K_p, which is considered to be measure of solar-terrestrial disturbances. Twenty-seven-day solar rotation periodicity was observed by the micropulsation activity. Occasionally there were rather sudden appearances c oscillations of great amplitude. The beginnin times generally coincided with sudden-commencement magnetic storms.

Daily observations—Of the 4808 hours sampled in the period from March through September 1958, 60 per cent had micropulsations. Th median amplitude in the north-south directio was 0.02 gamma. The usual period was abou 22 sec, but occasionally, during large-amplitud disturbances or at night-time, periods as rapi as 5 sec were measured. Sample data fo moderate day-time activity and the unusu night-time activity are shown in Figure 1.

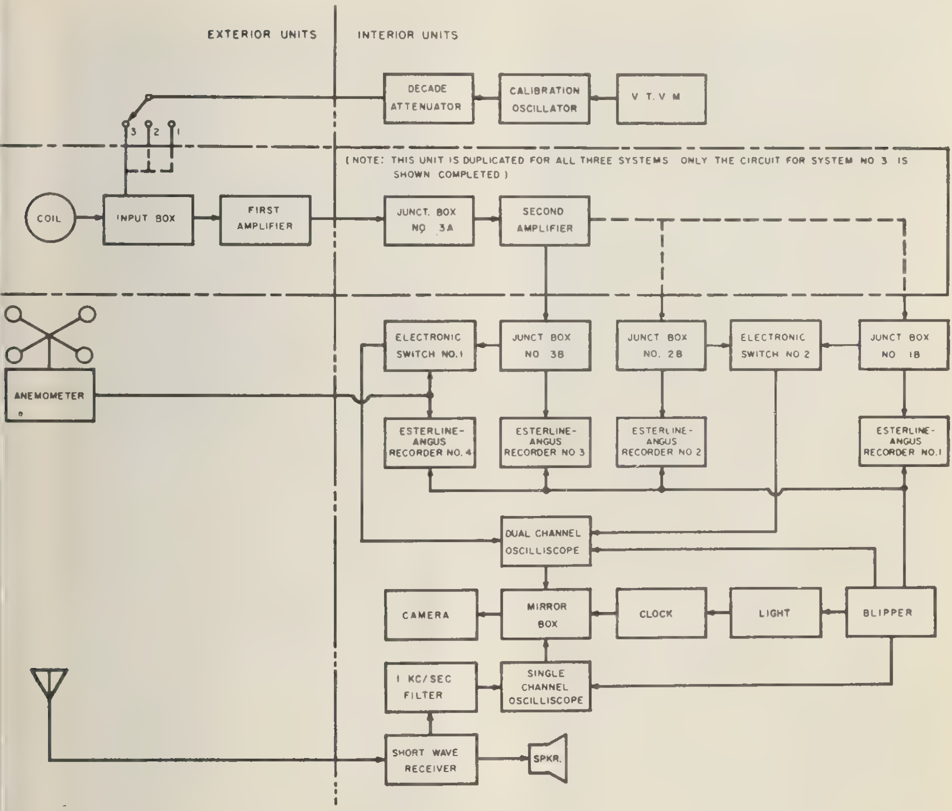


FIG. 3—Block diagram of instrument.

The average diurnal behavior of the rms agnitudes of magnetic flux density attained y the micropulsations in any 15-min interval re shown in Figure 2. The signals rise to axima of about ¼ gamma at 0945 and 1400 ours local mean time.

Occasionally three separate coil-antenna sysems were operated to obtain direction and otal field strength. At such times the signals ere displayed on an oscilloscope and the trace as recorded together with WWV time signal n slowly moving film. Figure 3 is a block diaram of the instrument. The frequency reonse for these recordings was extended to 20 ycles/sec. The micropulsations in the photoraph (Figure 4) indicate a north, east, and pward (and reverse) direction. The lesser mplitude is that of the vertical-axis antenna. he small sinusoidal trace is that of the anemometer recording a 3- to 5-miles/hr breeze. The rastered trace is the WWV time signal.

An average of a sample of the three-component flux density vectors gave $B_x = 0.24$, $B_y = 0.22$, and $B_{-z} = 0.07\gamma$. The north-axis flux density measurements should be multiplied by approximately 1.4 to indicate total field changes.

From the loop antenna data and earlier measurements made with a vertical antenna [*Campbell*, 1959; *Deal*, 1956] it was apparent that the average very low frequency natural electromagnetic spectrum has a transition from sferics to geomagnetics between 2.0 and 0.2 cycle/sec. This minimum in natural magnetic flux density is shown in Figure 5 for the daylight hours in the middle latitudes, far from active world thunderstorm centers and during a moderately disturbed (magnetically) day.

It has been suggested [*Troyickaya*, 1955]

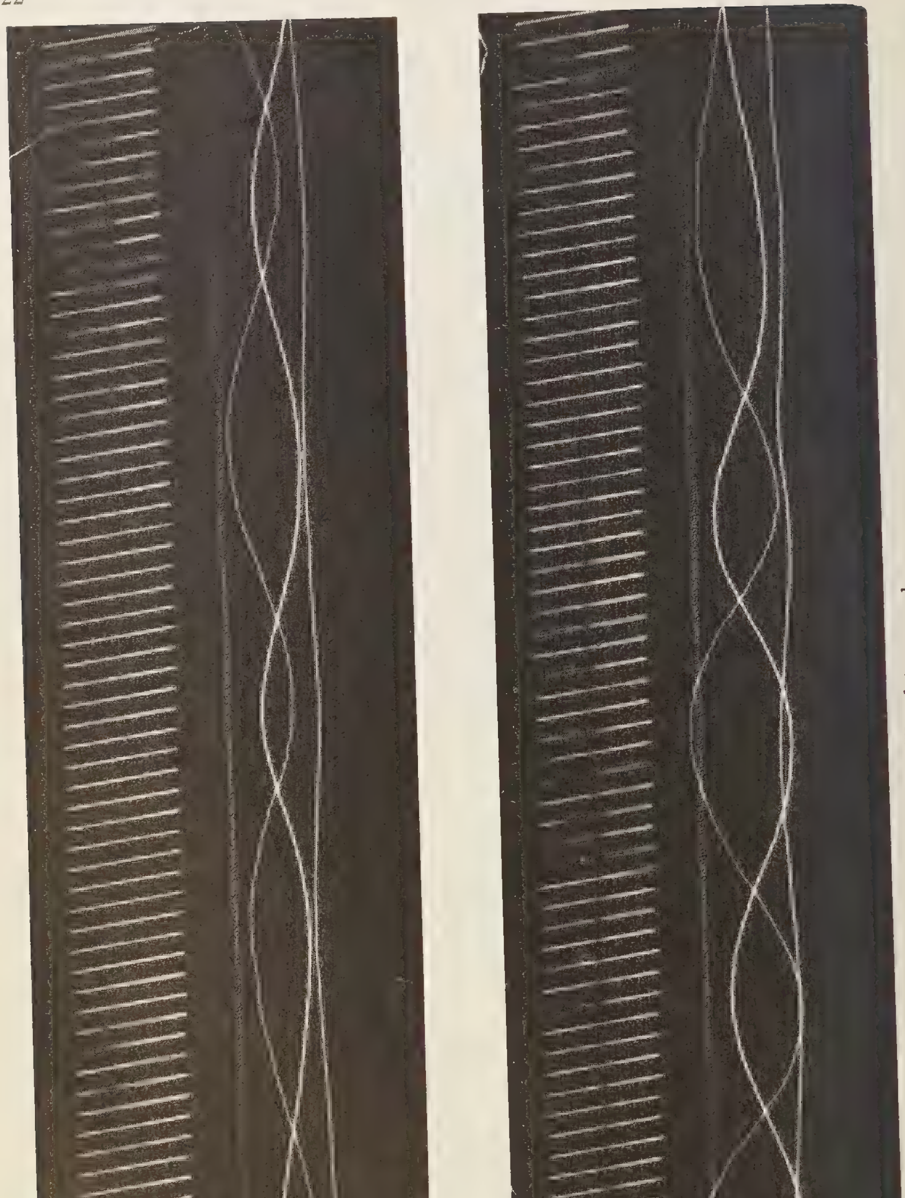


FIG. 4—Film data example.

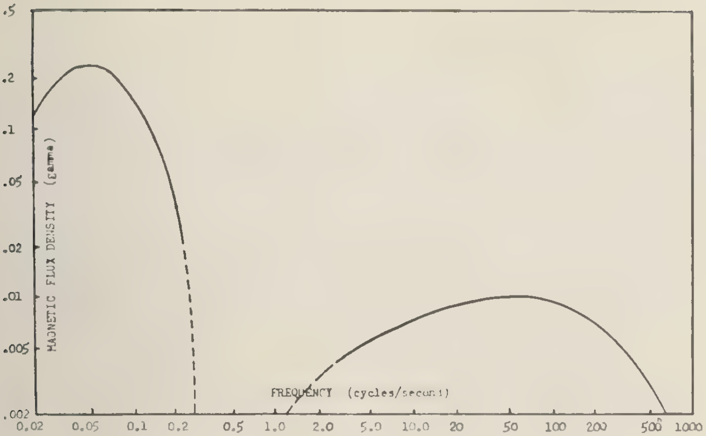


FIG. 5—Natural low-frequency electromagnetic spectra in California.

ıat the diurnal maxima of micropulsations ave a world time dependence perhaps tied to ıagnetic pole noon hours. In Figure 6 the data rom California, Germany [*Angenheister*, 1954], ıd Russia [*Troyickaya*, 1955] are listed under reenwich and local mean times. Note that the ırlier maxima are the larger ones for Caliırnia and Germany. A local time diurnal patrn is evident.

To illustrate the concurrent variation of miopulsations with F-layer parameters, I_f values[3] om White Sands, N. M., were plotted for the mple month of April 1958, along with the anetary daily average range index, A_p, and the ıurly micropulsation index,[4] M_h. Figure 7 shows the corresponding quiet and active occasions.

To illustrate the daytime correlation with the planetary 3-hour-range index, K_p was averaged for the 3-hour Greenwich intervals, 6, 7, 8, and 1 (hours of daylight in California) and the daily values were plotted versus M_d. Figure 8 shows the distribution.

The 27-day solar rotation cycle is apparent

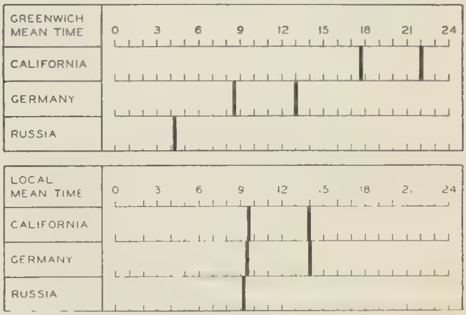


FIG. 6—Daily occurrence of micropulsations with periods of 5 to 30 sec.

[3] An idex of ionospheric F-layer variations was mposed in this manner: I_f is called a daily ytime index of ionospheric F-layer disturbances ual to one-third the sum of (a) the sum of the urly values of f_{min} for the day divided by the m of the hourly values of f_{min} for the monthly erage, (b) the sum of the hourly values of $h'F$ the day divided by the sum of the hourly lues of $h'F$ for the monthly average, and (c) e sum of the hourly values of f_oF_2 for the ınthly average divided by the sum of the urly value of f_oF_2 for the day. Only values ım 0700 to 2000 hours local mean time were ken.

[4] Hourly and daily micropulsation indices were fined to facilitate the statistical analysis of the ta. To each 15-min interval, named for the time the beginning of each interval, was assigned a lue equal to the maximum rms magnetic flux density in gammas attained by the oscillations in that interval. M_h was an hourly index equal to the average of the readable 15-min values following the selected hour. M_d was the average of the four largest M_h values for any one day. M_h and M_d values for March through September 1958 are available [*Campbell*, 1959].

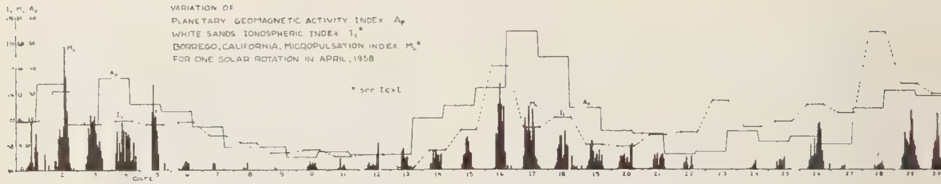


Fig. 7—April variation of A_p, I_p, and M_h.

in Figure 9, a 7-month record of M_d values.

Storms—Occasionally there appeared quite sudden increases in pulsation amplitude which were followed for several hours by higher activity than would have been expected for that day. This anomalous behavior was called a micropulsation storm. A typical one is shown in Figure 10.

In the 7 months of observation, 19 storms were studied (Table 1). Decision about what amplitude constituted a storm depended on the regular daily amplitude expected for the da Questionable storms were excluded from grou analysis.

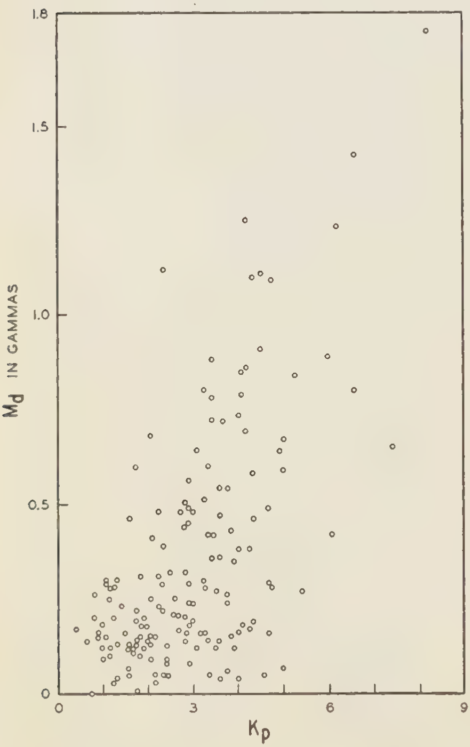


Fig. 8—Relation of micropulsation and disturbance indices.

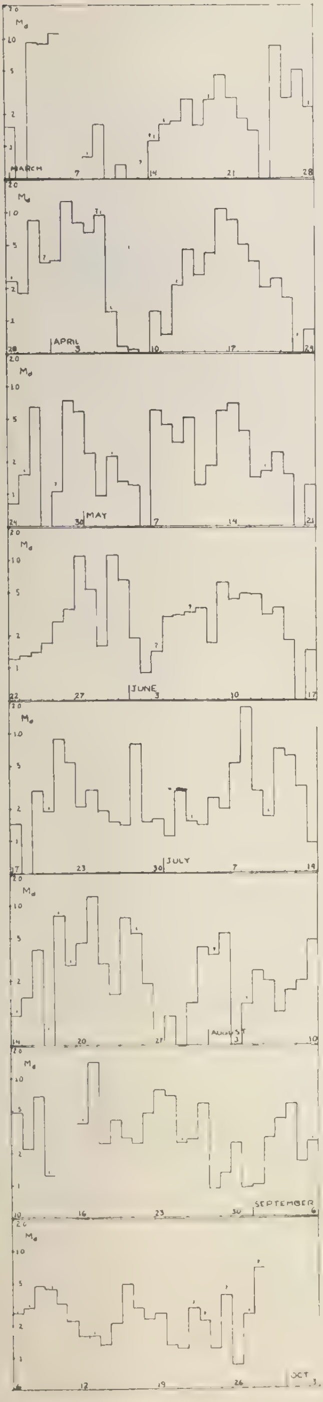
Table 1—*Micropulsation storms studied*

No.	Month	Day	GMT	Peak ampli-tude, gamma
	Local Nighttime			
1	July	8	0757	2.14
2	Aug.	17	0622	1.11
3	Aug.	26	0658	1.31
4	Sept.	25	0418	0.54
5	Sept.	25	0704	0.35
6	Sept.	25	0948	1.24
	Local Daytime			
1	March	25	1540	0.70
2	April	30	1700	1.40
3	May	26	1415	1.51
4	May	31	1652	2.76
5	June	14	1829	1.16
6	June	22	2030	1.96
7	June	28	1744	1.46
8	July	2	1647	0.80
9	July	21	1638	2.21
10	July	31	1532	0.71
11	July	31	1718	1.47
12	Sept.	25	1905	1.37
13	Sept.	28	2144	2.87

Of the micropulsation storms reported, t amplitude of the commencement had an avera value of 1.37 gammas. The storms were usual of shorter-period oscillations at the beginnin The occurrence distribution of micropulsati storms seemed to have maxima at 0700 a 1645 hours GMT. The average course of t storms was plotted for 5-min interval amp- tudes, as percentage of commencement peak, a storm time scale (Figure 11). The appea- ance of a large second maximum at 65 min indeed interesting. Storms occurred in the active part of the 27-day micropulsation cycle and were usually coincident with sudden-commencement magnetic storms.

Concluding remarks—The local diurnal fluctuation of micropulsations with periods of 5 to 30 sec is reminiscent of the diurnal fluctuation of mean electron densities at specific heights in the ionosphere and of diurnal fluctuation of ionospheric layer thickness [*Schmerling*, 1958]. The solar control of micropulsation amplitude was indicated by the M_h correspondence with magnetic disturbance indices, the *F*-layer fluctuation, and the 27-day cyclic behavior.

The micropulsation storms are a part of the world magnetic storm system and should be included in any proposed theory [*Kato and Watanabe,* 1958]. The regular local diurnal increase in amplitudes is superposed on the storms, which probably have a world time dependence. The 65-min secondary maximum is more likely a characteristic of the micropulsation storm itself than a part of the original excitation.



G. 9—Twenty-seven-day cycle of micropulsations.

References

Angenheister, G., Registrierung erdmagnetischer Pulsationen Gottingen, 1952/53, *Gerlands Beitr. Geophys., 64,* 108–132, 1954.

Campbell, W. H., A study of micropulsations in the earth's magnetic field, *Inst. Geophys., Univ. Calif. Los Angeles Sci. Rept. 1,* Nonr 233(47), 1–138, 1959.

Deal, O. E., Observation of very low frequency electromagnetic signals of natural origin, Ph.D. Thesis, University of California at Los Angeles, 1956.

Holmberg, E. R. R., Rapid periodic fluctuations in the geomagnetic field, I, *Monthly Notices Roy. Astron. Soc., Geophys. Suppl., 6,* 467–481, 1953.

Kato, Y., and T. Watanabe, A survey of observational knowledge of the geomagnetic pulsation, *Sci. Repts. Tôhoku Imp. Univ., ser. 5, Geophys., 8* (3), 157–185, 1957.

Kato, Y., and T. Watanabe, Studies on Geomagnetic Storm in Relation to Geomagnetic Pulsation, *J. Geophys. Research, 63,* 741–756, 1958.

Schmerling, E. R., Ionospheric electron density-height profiles, *I. G. Y. Bull. 15,* 1958, *Trans. Am. Geophys. Union, 39,* 1018–1021, 1958.

Stewart, B., On the great magnetic disturbance of 28 August to 7 September 1859, *Phil. Trans. Roy. Soc. London, 151,* 423, 1861.

Sucksdorff, E., Giant pulsations recorded at Sodankyla during 1914–1938, *Terrestrial Magnetism and Atmospheric Elec., 44,* 157–170, 1939.

Troyickaya, V. A., Earth currents, *Priroda, 5,* 81 85, 1955.

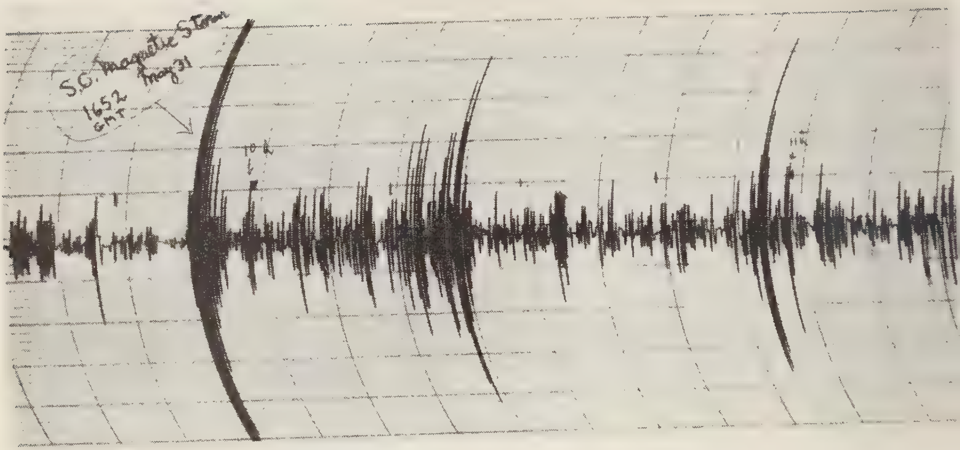


Fig. 10—Example of micropulsation storm.

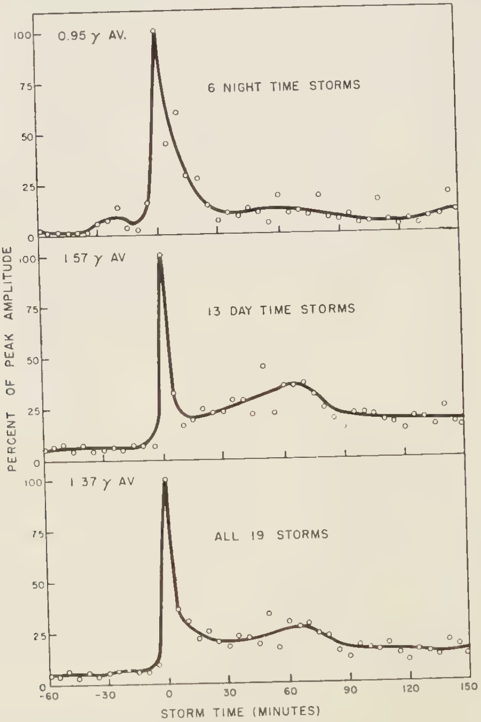


Fig. 11—Average micropulsation storms.

The Relationship between Geomagnetic Variations and the Circulation at 100 Mb[1]

Julius London, Irwin Ruff, and Leo J. Tick

Department of Meteorology and Oceanography, and Research Division
College of Engineering, New York University, New York 53, N. Y.

Abstract—Statistical methods are used to study the relationship, over a five-year period, between geomagnetic storms and the height gradients of the 100-mb surface over the United States. In the case of the superposed epoch analysis, the parameters studied were variations from population values of the mean height gradients and the standard deviation of the height gradients for the period 5 days before to 15 days following the geomagnetic storm. No pattern was found in the tested statistical significance at the 5 per cent and 1 per cent levels. The geomagnetic and height gradient spectra were constructed from the two time series and the coherency between the two series was computed. The coherency was found to be very small for all periods from 4 to 60 days. The conclusion is drawn that there is no obvious relationship between the two sets of data.

Introduction—Many attempts have been nade in the past to connect anomalous solar acivity, such as solar flares, sun spottedness, and orpuscular emissions, with changes in the arth's atmosphere. These relations have been airly well established for such high atmospheric henomena as aurorae and ionospheric and eomagnetic fluctuations. In the case of the irculation of the troposphere and lower stratophere, however, no definite relationship has een shown.

The principal variations of solar electromagetic energy are found in the very short (far ltraviolet and X-ray) or in the very long radio) wavelengths. Although the percentage hanges in these spectral regions may be very reat, the total energy enhancement is small ompared with the energy in the visible portion f the solar spectrum.

Large variations of solar energy output also ccur in the form of periodic and sporadic ejecions of charged particles, principally ionized ydrogen. The total energy involved in such rpuscular radiation is small compared with e electromagnetic radiation in the visible range. owever, the corpuscular radiation is of special interest due to the fact that, since the charged particles are deflected by the earth's magnetic field, the particles enter the atmosphere principally at high latitudes (the auroral zones). The incoming particles produce two well-known effects in the atmosphere: the aurorae and the disturbances of the geomagnetic field.

It will not be argued here that it is impossible for the circulation patterns of the lower atmosphere to affect the ionospheric winds in such a manner as to cause some fluctuations in the geomagnetic field, as, for instance, discussed by *Wulf and Hodges* [1950]. There is no doubt, however, that the large fluctuations in geomagnetic activity result directly from anomalous solar activity.

Fluctuations of the geomagnetic field are readily measurable in a quantitative manner. Also, measurements of the geomagnetic field may be made continuously and are not interrupted by extraneous factors, such as cloudiness. It is therefore logical, in studying atmospheric effects believed due to solar corpuscular radiation, to use geomagnetic activity as a measure of the intensity of this radiation. Also, it seems reasonable to expect that some other effects besides aurorae and geomagnetic disturbances should be produced by high-speed particles, since only a fraction of their kinetic energy is used in producing these phenomena [see, for example, *Sugiura and others,* 1952]. The ques-

[1] This research was supported by the Geophysics esearch Directorate, Air Force Cambridge Rearch Center, Bedford, Mass., under Contract AF)(604)-1738.

tion remains whether this excess energy, either directly or indirectly, manages to produce large-scale changes in the circulation of the troposphere and lower stratosphere.

Since the incoming solar particles represent a flux of energy entering the atmosphere from above, it seems plausible that one should, in seeking an influence on the circulation of the atmosphere, examine the highest level for which synoptic data are available for a fairly long period. At present this appears to be the 100-mb level (16 km).

The results of *Shapiro* [1956] and *Macdonald and Roberts* [1958] suggest that it is the large-scale feature of the atmospheric circulation pattern that might be associated with geomagnetic fluctuations. In the present study, the large-scale pattern was characterized by the 100-mb zonal circulation over the central United States.

Geomagnetic data—There are several methods of expressing the world-wide variation of geomagnetism. One of the most precise and widely used of these indices is K_p, the three-hour geomagnetic range index. K_p is so designed that there is an exponential relationship between it and the range.

Since the large increases of geomagnetic activity vary considerably both in magnitude of the peak value and in the abruptness of the rise, there is some question as to what exactly constitutes a geomagnetic storm. For the purpose of investigating a proposed solar-atmospheric relationship, however, it might be expected that the important factor in producing an atmospheric effect would be an event resulting in a sudden (1- or 2-day), large increase of geomagnetic activity. Thus a 'storm' may be defined by specifying a magnitude and rate of increase of K_p. (It is here assumed that it is not the actual value of the geomagnetic activity, but rather an impulse-like increase of sufficient magnitude that is important in producing an atmospheric effect.) Presumably the larger the impulse, the stronger will be the atmospheric reaction. It was therefore decided to define two geomagnetic storm intensities, one characterizing moderate storms and the other—a more restrictive definition—corresponding to strong geomagnetic storms.

The definitions used are

a. *Moderate*—A steady rise of the three-day running mean of K_p equal to or greater than 1 in two days or less.

b. *Strong*—A steady rise of the three-da running mean of K_p equal to or greater than 1 in two days or less.

The above storm definitions follow approxi mately from changes in the frequency distribu tion of the direction and magnitude of stead rises of K_p for the period studied. The use c three-day running means of the daily sums c K_p has the advantages of smoothing out th small daily fluctuations while retaining larg increases. There is some disadvantage in tha the peaks of the curves are somewhat flattene and may be displaced by one day.

In each case a particular day was chosen t indicate the time of the geomagnetic storn This representative day was somewhat arbi trarily chosen as the end of the defined perio of increase, with the possibility of choosing day slightly beyond this period if the rise con tinued at a sufficient rate.

Meteorological data—The upper-level circu lation patterns were approximated by takin average height differences with latitude of th 100-mb surface over the United States. Th region has the advantage of relatively good dat coverage at that level.

The basic meteorological data used were th reported 0300Z 100-mb heights at Great Fall Mont.; International Falls, Minn.; Moosone Ont.; Grand Junction, Colo.; Columbus, Mo Pittsburgh, Pa.; El Paso, Tex.; Lake Charle La.; and Charleston, S. C. for the period 195 1955. These stations, shown in Fig. 1, we chosen to approximate a grid consisting of th

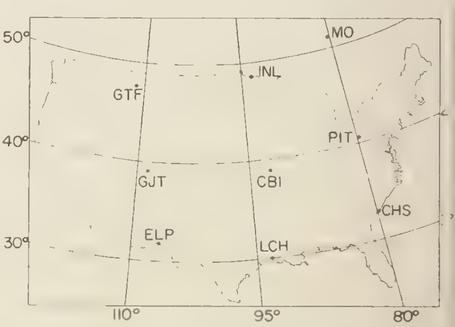


Fig. 1—Network of stations (U. S.) used in co puting the 100-mb height gradients.

intersections of parallels 30°N, 40°N, and 50°N with meridians 80°W, 95°W, and 110°W. Thus the grid covers the central three-fourths of the United States. Daily means of the heights along each of these parallels were taken, and daily differences were then obtained between the means along adjacent parallels. These height differences are proportional to the mean u (zonal) component of the geostrophic wind at 100-mb.

Method—The primary statistical method of analysis was that of superposed epochs, a method which can be used to determine whether or not two time variables are related [see, for instance, *Panofsky and Brier*, 1958]. If the two variables are related, even with a lag, it will be revealed as a change in the values of the statistic examined. Spectra and cross-spectra for the two sets of data (geomagnetic and meteorological) were also computed and the coherency of the pair of series was determined.

For the method of superposed epochs, the key time was taken to be a day of geomagnetic storm. The seasonally corrected three-day running means of the mean 100-mb height gradients were tabulated for all key days, and for the days from 5 days before the key day to 15 days after, and the means and standard deviations of the 21 resultant distributions were calculated. The analysis covered the full five-year period 1951–1955.

The purpose of using the standard deviations

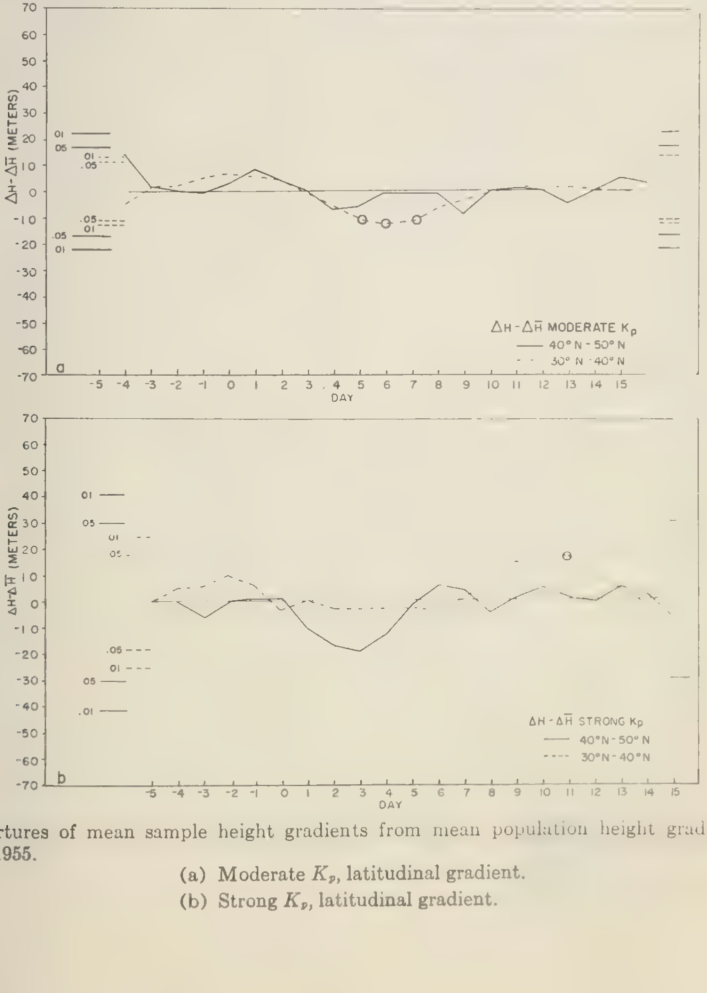


FIG. 2—Departures of mean sample height gradients from mean population height gradients for the period 1951–1955.

(a) Moderate K_p, latitudinal gradient.

(b) Strong K_p, latitudinal gradient.

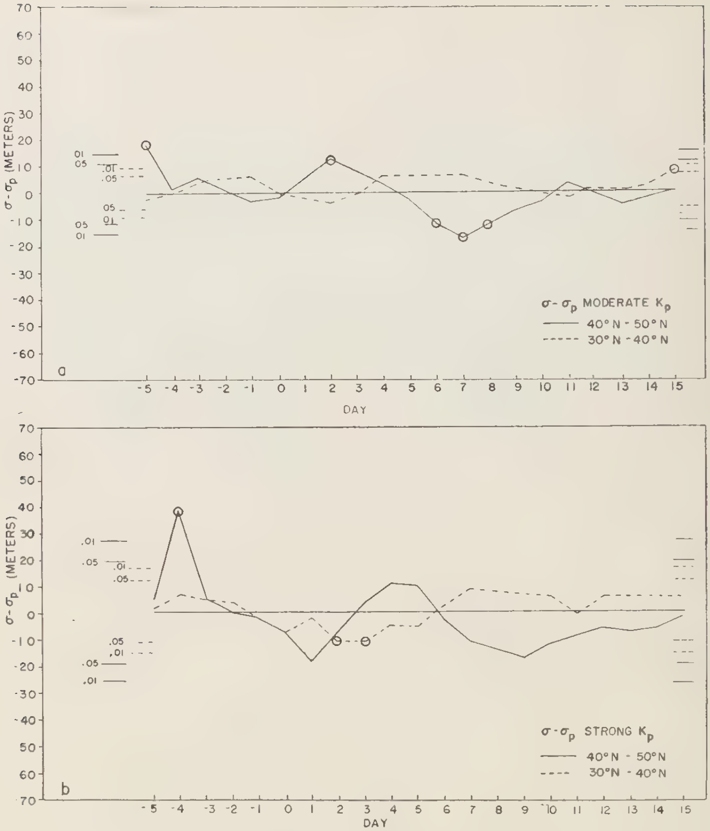


FIG. 3—Departures of sample standard deviation of height gradients from population standard deviations of height gradients for the period 1951–1955.

(a) Moderate K_p, latitudinal gradient.

(b) Strong K_p, latitudinal gradient.

was to investigate whether a relationship between geomagnetic storms and 100-mb height gradients might be present even though the response was not always in the same direction. For example, it is possible that the individual mean height gradient does respond to a magnetic storm but, depending on the initial pattern, sometimes increases and sometimes decreases. In that case the curve of the mean height gradients might not show any major variation. It would be expected, however, that the curve of the standard deviations of the 21 distributions would indicate such a response.

The mean or standard deviation of the mean height gradients for any one of the days −5 to 15 is taken to be the mean or standard deviatio of a sample taken from a population which consists of the values for the entire period considered. Comparison of any one of the sampl means or standard deviations with the corresponding population value determines whethe or not it differs significantly from the populatio value (that is, whether or not it is significant) It should be noted that the significance of eac of the sample parameters must be determine separately, since the samples are not independent. This is due to the persistence of the value of the height gradients from day to day, whic persistence is accentuated by the use of three day running means.

The means were tested for significance by use of the variate

$$t = \frac{(\bar{x} - \mu)\sqrt{n}}{\sigma}$$

where $\bar{x}$ is the sample mean, μ the population mean, σ the population standard deviation, and n the sample size. A normal distribution of the means was assumed.

The F test was used to determine the significance of the standard deviations. Although the mean height differences are not normally distributed (the F test is strictly valid only for normally distributed parameters), the frequency distribution of height differences resembles a normal distribution sufficiently well, and the F test is sufficiently flexible for this test to be used.

The significance levels considered in this study were 0.05 and 0.01.

Results and discussion—The results of the superposed epoch analyses are summarized in Figures 2 and 3. These present graphs of the deviations of the means and standard deviations from the corresponding population parameter as a function of the time preceding and following key days. The ordinates (labeled simply as meters) represent height gradients in meters per 10° of latitude. Thus, we are here essentially concerned with the mean and variance of the 100-mb zonal winds before and after geomagnetic storms.

In the analysis shown in Figures 2 and 3 it is seen that there are 4 points for the means and 9 points for the standard deviations that are significant at the 5 per cent level. The number of points significant at the 1 per cent level are of course much fewer, being only 1 and 3 for the means and standard deviations, respectively. The significant points are indicated by circles on the curves in Figures 2 and 3. While the number of significant points for the means are only what would be expected by chance (there being 21 × 2 × 2 = 84 individual days considered), the number of significant points for the standard deviations far exceeds expectation. Examination of the curves in Figure 3, however, raises certain pertinent questions concerning the physical interpretation of these results. Because of the definition of the K_p storm day, significant points occurring on or before day −3 have no meaning since the key day is selected in such a way as to represent an impulse evident in the geomagnetic field, which impulse will at some subsequent time cause a change in the 100-mb circulation patterns. Also, it is evident that a significant day on one set of curves is, in general, quite different from the significant days on the other curves in Figure 3. The phase reversal shown between the solid and dashed curves in Figures 3a-b is certainly fortuitous. Of particular importance is the fact that there is no marked increase of the standard deviation in the time interval of 10 to 14 days following the geomagnetic storm as might be expected from the results of *Shapiro* [1956] and *London* [1956].

The data were also divided into winter (Dec., Jan., Feb.) and summer (June, July, Aug.) groups to determine whether or not there is a seasonal effect which is hidden by using all the data together. As would be expected, the winter variance was much larger than the summer variance but on no day was the standard deviation significant even at the 5 per cent level.

Further analyses were made using A_p, the average planetary amplitude, as the geomagnetic index and redesignating the geomagnetic storm day as the day of the beginning of a sharp rise in the geomagnetic index. In addition, the meteorological data were extended to cover a portion of western Europe for the period 1952–1953. The results, discussed by *London and others* [1959], indicate even less significance of the meteorological variations than those shown above.

Spectrum analysis—Since the method of superposed epoch may involve some interaction of the meteorological variables during two or more successive geomagnetic storms, and certainly involves some arbitrariness as to the choice of the impulse day (key day), it was decided to make a spectral analysis of the time series. In this way all the data were used, in the time order in which they arose. The two time series used cover the 5-year period 1951–1955 for the daily values of K_p and the daily values of the mean 100-mb height differences between latitudes 30°N and 50°N as discussed above.

The statistic computed was the coherency between the two series, which is defined as

$$\text{coherency} = \frac{(\text{co-spectrum})^2 + (\text{quadrature spectrum})^2}{(\text{spectrum}_1)(\text{spectrum}_2)}$$

The coherency is, of course, a function of frequency and is the square of the correlation coefficient between the same frequency component of each series. As is true of all correlation coefficients, this is a measure of the intensity of the linear relationship of random variables. The results are shown in Figures 4 and 5. Figure 4 gives the spectra of the two series. The spectrum for K_p shows peak energy at the period of about 25 to 30 days, as is to be expected. Other, less important peaks are present at about 13 to 15 days, 8 to 9 days, etc. The spectrum of 100-mb height gradients (mean zonal wind over U. S.) is very similar to the spectra of 500-mb zonal indices for these latitudes [*Panofsky*, 1956] in that it shows a slight peak at about 25 days and the typical pattern for large-scale meteorological phenomena; that is, mostly long-period energy.

As can be seen from Figures 4 and 5, although the two spectra show some slight similarity at the longer periods, the coherency between the two series is extremely small and represents almost exactly the type of pattern one would expect between two completely uncorrelated series. (The average value of the statistic under zero coherency is, from sampling experiments, approximately 0.08 for the degrees of freedom used.)

The maximum coherency found corresponds to a correlation coefficient of about 0.3 and occurs at lags representing periods of abou 12, 7, and 5 days, where, as can be seen in Fig ure 4, the power spectra show relatively litt energy. There is currently no exact significanc test available; however, the comparison wit sampling experiment is an approximate tes There is certainly no obvious evidence of an linear relationship between the two series.

Conclusions—The conclusion to be drawn therefore that no relationship has been show in our analyses between geomagnetic storm and 100-mb height gradients. It is certainly po sible that some actual relationship does exi which was not revealed. However, if there we any definite response of the 100-mb circulatic patterns to the solar corpuscular radiation th causes the geomagnetic storms, it would pr sumably show up in these analyses. Even if th response were not always the same but d pended on the initial state of the atmospher the corpuscular radiation acting only as a tri ger, the analysis of the standard deviatio should reveal the presence of a response. Th any direct effect must be negligibly sma causing no distinguishable variation in the larg scale circulation patterns.

It may be that longer time series than we taken here are necessary to show the anom lous atmospheric reaction to solar influences, it exists at all. A recent study of the vertic

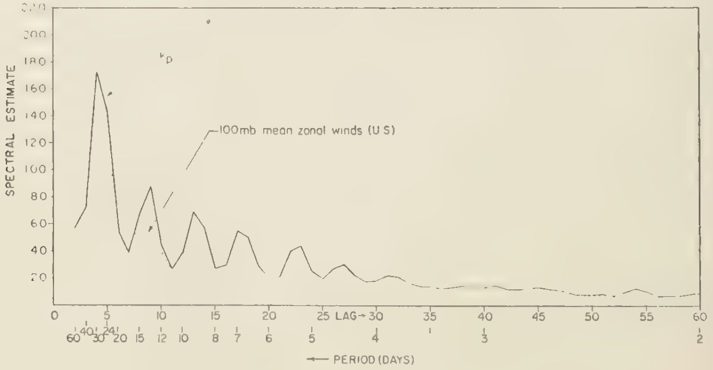


Fig. 4—Spectral estimates of K_p and 100-mb mean zonal winds (U. S.).

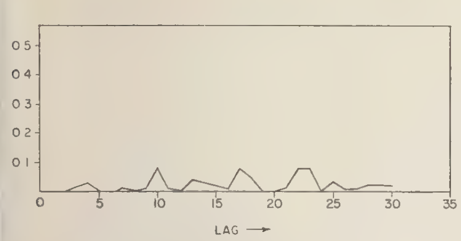


FIG. 5—Coherency between K_p and $\bar{U}$.

ropagation of hydrodynamical pulses through he mesosphere [*Ooyama,* 1958] indicates, howver, that such a response would, at best, be xtremely small.

REFERENCES

ONDON, J., Solar eruptions and the weather, *Trans. N. Y. Acad. Sci., 19,* 138–146, 1956.

ONDON, J., I. RUFF, AND L. J. TICK, The relation between geomagnetic variations and the circulation at 100 mb, *Sci. Rept. 8, Contract AF 19(604)-1738,* Research Division, College of Engineering, New York University, 1959.

ACDONALD, N. J., AND W. O. ROBERTS, The relationship of 300 mb circulation change to geomagnetic disturbances from 1952 to 1958, *Tech. Rept. 6, Inst. for Solar-Terrest. Research,* High Altitude Observatory of the University of Colorado, 1958.

OOYAMA, K., On the vertical propagation of a disturbance through the mesosphere, *Sci. Rept. 6, Contract AF 19(604)-1738,* Research Division, College of Engineering, New York University, 1958.

PANOFSKY, H. A., Cross spectrum analysis of hemispheric 500 mb zonal indices at latitude 25°N and 60°N, *Research Rept. on Task 27, Project AROWA,* U. S. Navy, 1956.

PANOFSKY, H. A., AND G. W. BRIER, *Some Application of Statistics to Meteorology,* Penna. State Univ., 224 pp., 1958.

SHAPIRO, R., Further evidence of a solar-weather effect, *J. Meteorol., 13,* 335–340, 1956.

SUGIURA, M., M. TAZIMA, AND T. NAGATA, Anomalous ionization in the upper atmosphere over the auroral zone during magnetic storms, *Rept. Ionosphere Research Japan, 6,* 147–154, 1952.

WULF, O. R., AND M. W. HODGES, On the relation between variations in the earth's magnetic field and variations of the large-scale atmospheric circulation, *J. Geophys. Research, 55,* 1–20, 1950.

(Manuscript received June 8, 1959.)

A Vertical Cross Section Through the 'Polar-Night' Jet Stream

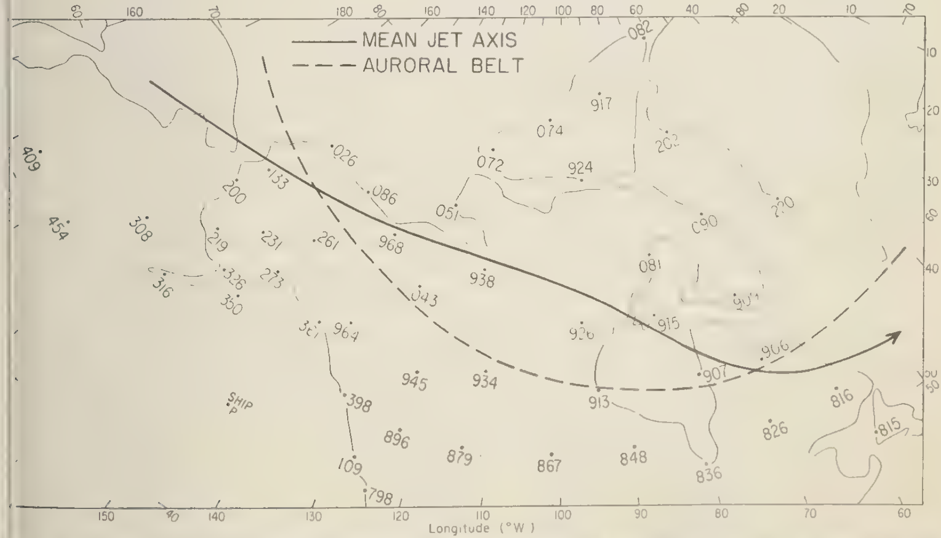
T. N. Krishnamurti

Department of Meteorology, University of Chicago
Chicago, Illinois

Abstract—Because of the great altitude of the core of the 'polar-night' jet stream, only isolated rawinsonde observations have penetrated the core, and this scarcity of data renders the construction of synoptic cross sections difficult. For a more definitive determination of the structure of this current, all soundings of the North American Arctic were combined into one cross section for a four-day period when the jet stream was in relatively steady state. It turned out that the core was located at a height of 26 km and had a speed of 135 knots. Below this altitude the atmosphere was isothermal in the mean; above it, temperatures increased upward.

Cross sections were constructed for the wind components parallel and normal to the jet axis, temperature, potential temperature, and absolute and potential vorticity. Comparison was also made between observed and geostrophic wind speeds; a high correlation was found to exist.

Introduction—From soundings taken at ›isko, Sweden (68°N), during the 1920's, *ılmén* [1934] concluded that the hypothesis a warm stratosphere over the Arctic in win- ɪ: was not tenable, and that, on the contrary, ry cold air was located there at high altitudes. ter observations made in the Arctic and the Antarctic [*Court*, 1945] confirmed the very low (−80°C) midwinter temperatures in the polar stratospheres; in accord with Palmén's findings, they also revealed that midsummer temperatures are very warm. The former may result from prolonged cooling of the ozone layer at low solar altitudes or during darkness, the latter



Fɪɢ. 1—Outline of North American continent, mean position of axis of polar-night jet stream for the ːiod from December 31, 1957, to January 3, 1958, and average position of center of belt of aurora; tions used in investigation are marked with dots and numbers; corresponding station names are ınd in Table 1.

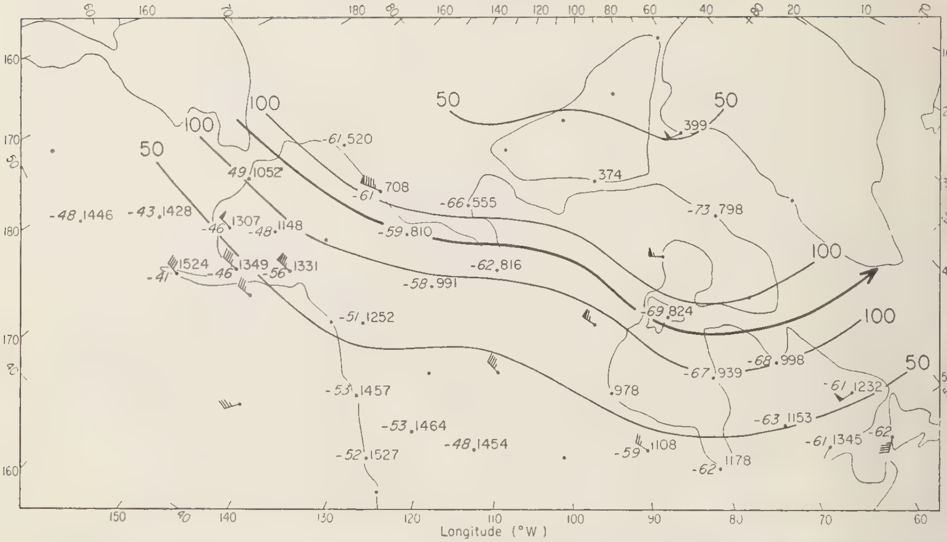


FIG. 2—50-mb isotachs (knots), December 31, 1957, 12 Z. Winds at 00 Z have been added where n 12 Z report was available; temperature in °C and height of 50-mb surface in meters, departure fro 19,000 m; heavy line denotes approximate position of jet axis.

from permanent insolation during summer. From the contributions on geostrophic computations the conclusion was drawn that a wintertime maximum of westerly wind must be located over the Arctic and the Antarctic at altitudes barely accessible to observation until recent years [*Heastic,* 1955; *Hess,* 1948; *Kochanski,* 1955; *Loewe and Radok,* 1950; *McIntyre,* 1955].

Coincident with advances in rawinsonde techniques, vertical cross sections have been drawn through the Arctic-night temperature gradient in recent years with isotach analysis based on the wind observations [*Godson and Lee,* 1957, 1958]. These have confirmed the existence of a current with the core near 25 mb or higher and with lateral velocity gradients typical of jet streams. On account of the great altitude of the core, however, these sections usually contain only a single ascent with winds penetrating into the core, so that definitive analysis is difficult. The purpose of this study has been to compute a cross section from a large number of observations by combining all stations of the North American Arctic for a period of several days during which fluctuations of the current were not excessive and enough data existed to make it possible to locate the jet axis approximatel All individual balloon ascents were to be com posited in a 'jet coordinate system'; that is, a data located at similar distances from the ax of the current were to be averaged.

Determination of axis—A survey of dai upper winds during the winter of 1957–195 indicated that the period from December 3 1957, to January 3, 1958, although not idea would serve the above purposes as well as an other period for which data were availabl Observations at 1200 Z were utilized primaril Nearly 200 upper-air soundings from 49 statio were available for analysis between latitudes 5(and 80°N, and longitudes 60° and 170° (Table 1, Fig. 1). The decrease in observati density with height, however, was very lar (Table 2); little more than 20 per cent of t ascents which passed through the 300-mb su face reached the 25-mb surface. This decay frequency may have introduced bias in the co putations to be presented, and the structure the current as determined here must therefc be regarded as only preliminary. In order base the calculations on the largest possil number of observations, winds at 50 and 25 r for 00 Z were included to supplement the no

TABLE 1—*List of upper-air stations*

Block number	Station number	Station name	Block number	Station number	Station name
70	026	Barrow	72	906	Fort Chimo
	086	Barter Is.		907	Port Harrison
	133	Kotzebue		909	Frobisher
	200	Nome		913	Churchill
	219	Bethel		915	Coral Harbor
	231	McGrath		917	Eureka
	261	Fairbanks		924	Resolute
	273	Anchorage		926	Baker Lake
	308	St. Paul Is.		934	Fort Smith
	316	Cold Bay		938	Coppermine
	326	King Salmon		945	Fort Nelson
	350	Kodiak		964	Whitehorse
	361	Yakutat		968	Aklavik
	398	Annette	74	043	Norman Wells
	409	Attu		051	Sachs Harbor
	454	Adak		072	Mould Bay
72	815	Stephenville		074	Isachsen
	816	Goose Bay		081	Hall Lake
	826	Nitchequon		082	Alert
	798	Tatoosh Is.		090	Clyde
	836	Moosonee		109	Port Hardy
	848	Trout Lake	SHIP	4YP	Phappa
	867	The Pas	04	202	Thule
	879	Edmonton		220	Egedesminde
	896	Prince Georg			

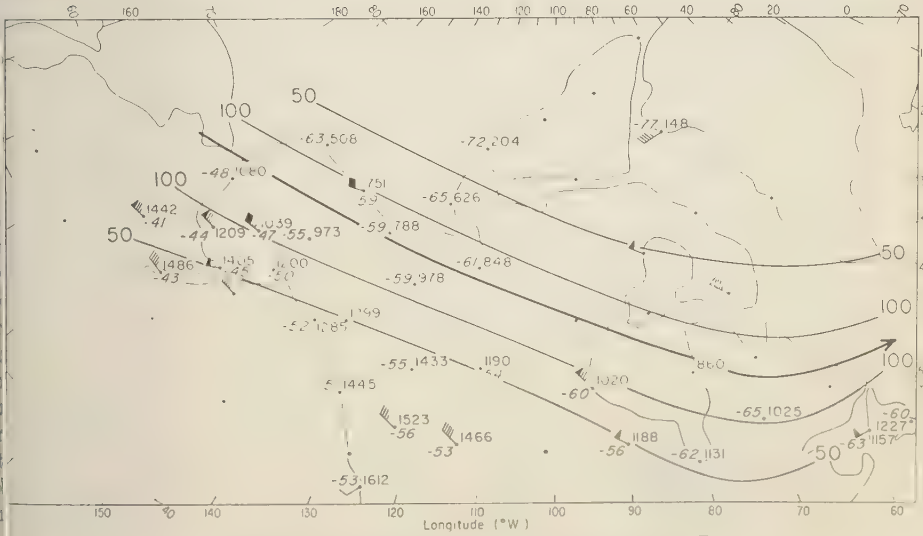


FIG. 3—50-mb isotachs (knots) for January 1, 1958, 12 Z.

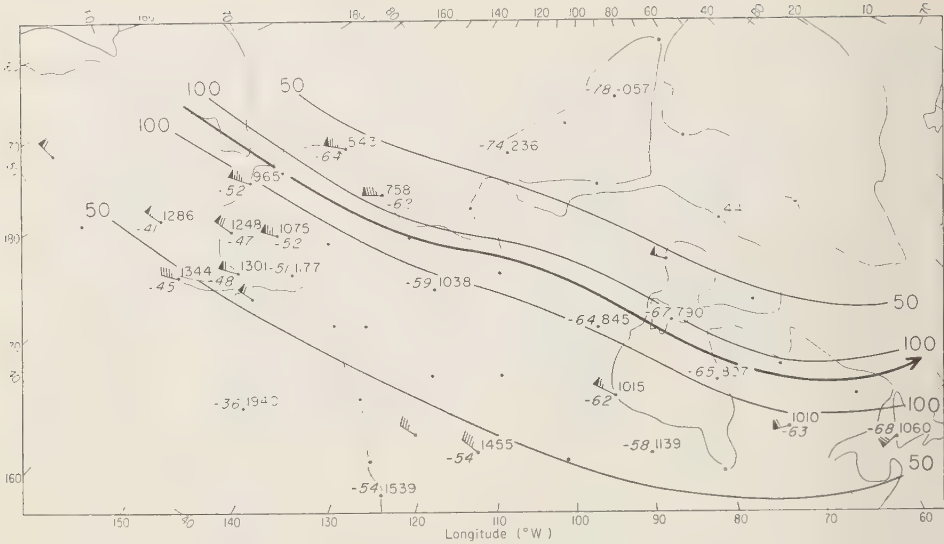


FIG. 4—50-mb isotach (knots) for January 2, 1958, 12 Z.

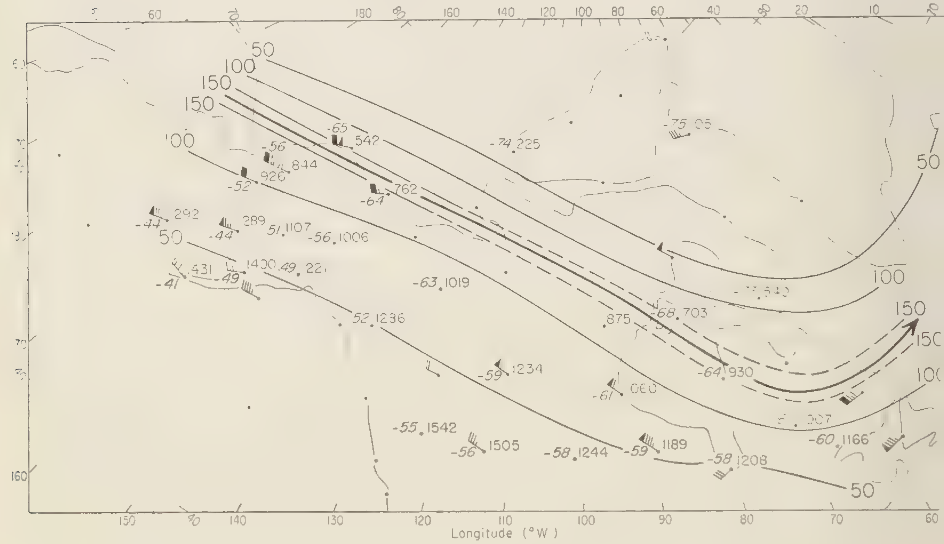


FIG. 5—50-mb isotachs (knots) for January 3, 1958, 12 Z.

data. Few additional data points, however, were gained by this procedure because winds changed little in the 12-hour intervals and the high ascents were made mostly at the same stations.

The first step consisted in locating the jet-stream axis on each of the four days, where the axis is defined as the line following the m mum wind speeds in an isobaric surface. Initi it was attempted to do this at 25 mb, but servations were too sparse. At 50 mb a l minimum of data was available. In drav the isotach charts (Figs. 2 to 5) continuity

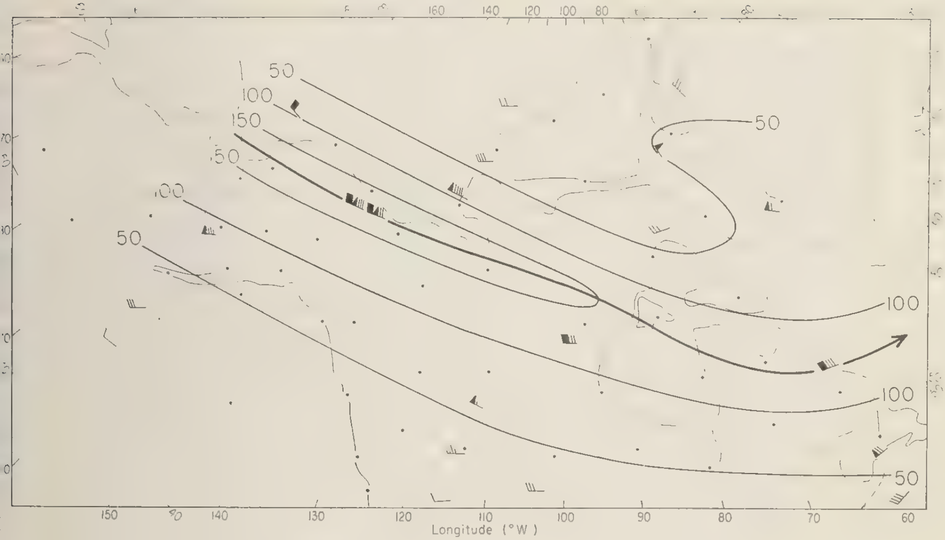


FIG. 6—25-mb isotachs (knots), with all winds of period combined.

sed; one could assume that the thermal wind quation holds for the current and apply it ualitatively, because the axes as suggested by he wind data were located in the zone of strong oleward temperature gradient. On account of ne well-known difficulties in computing the eight of constant-pressure surfaces in the tratosphere, calculations of geostropic winds ere not attempted on individual charts.

From Figures 2 to 5 the mean position of the xis for the four days was determined (Fig. 1). 'he mean latitude was about 65°N and the orientation of the axis followed the belt of aurora more closely than a latitude circle. The auroral belt referred to is the winter mean position (not for this period). As also noted by other authors [*Austin and Krawitz,* 1956; *Teweles,* 1958], the arctic jet stream frequently does not lie in an axially symmetric position with respect to the North Pole. This is a surprising phenomenon because the current exists far removed from surface influences and is not well correlated with tropospheric disturbances [*Austin and Krawitz,* 1956].

TABLE 2—*Total number of radiosonde and wind observations used in the study*

Pressure level, mb	Radiosonde	Wind	Percentage* of total
300	191	180	100
200	185	168	97
150	172	156	90
100	97	83	51
50	75	60	39
25	41	21	21
10	12	2	6
6	5	...	2

*Indicates the decay of observations relative to ie 300-mb level.

In Figure 6, all 25-mb winds for the four days have been plotted with respect to the mean axis position; that is, the winds do not appear at the sounding stations but at the appropriate locations in the coordinate system fixed in the axis. Winds above 50 knots covered a belt 20° of latitude wide; along the axis, highest speeds near 180 knots were located in the western ridge and minimum speeds near 120 knots in the trough over eastern Canada. This agrees with what the author has found for the subtropical jet stream of winter,[1] but it must be noted that

[1] T. N. Krishnamurti, "The Subtropical Jet Stream of Winter," unpublished technical report, Department of Meteorology, University of Chicago.

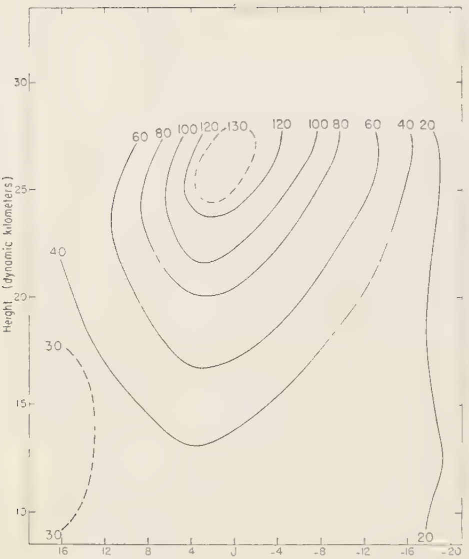


Fig. 7—Vertical cross section of wind component parallel to jet axis c_s (knots). Orientation of axis in Figure 1 was taken to define parallel and perpendicular wind components at all heights.

the distribution of speeds in Figure 6 may hav arisen merely from transient high-speed cente along the axis.

In angular measure, the distance from rid to trough was about 90° of longitude, suggestir a two-wave pattern if a circumpolar current ca be assumed [*Teweles*, 1958].

Wind cross sections—For the following con putations the normal distance between the ax position on each day and each observation w determined. Then all observations were groupe into class intervals of 4° latitude of normal di tance, where degrees-of-latitude is used as mea ure of length only and does not necessari denote latitudinal distance. All observations each class interval were then averaged, center on the axis; that is, observations from 2° on t left to 2° on the right were combined, from to 2° on either side, etc.

The winds were broken into components (c_n) in an *s*-*n* coordinate system where the *s*-a parallels the jet axis and *n* is taken normal to positive to the left. The c_s cross section (Fig. is spectacular, indicating a jet core with me speed of 135 knots at 26 dynamic km. As ge

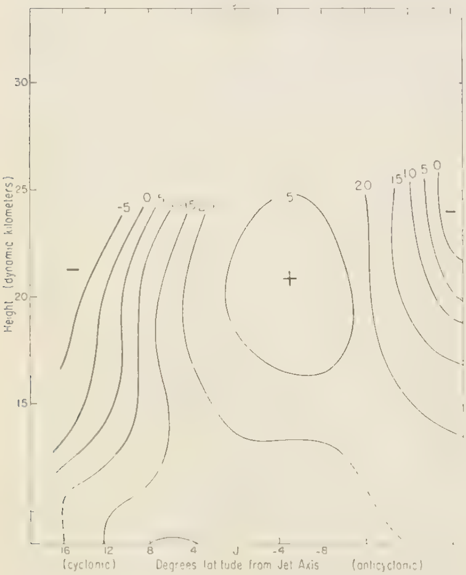


Fig. 8—Vertical cross section of wind component perpendicular to jet axis c_n (knots).

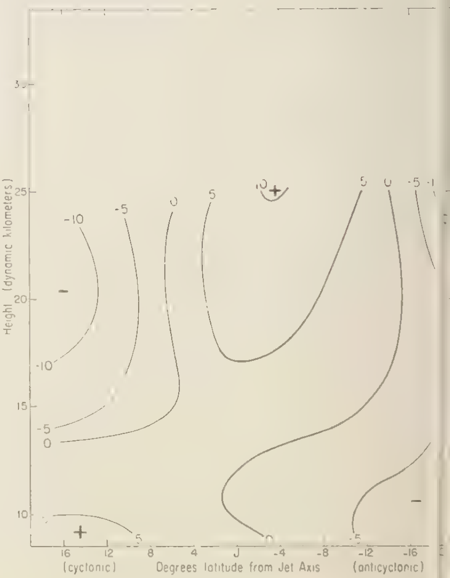


Fig. 9—Vertical cross section of mean ageos phic wind component perpendicular to jet axis (knots).

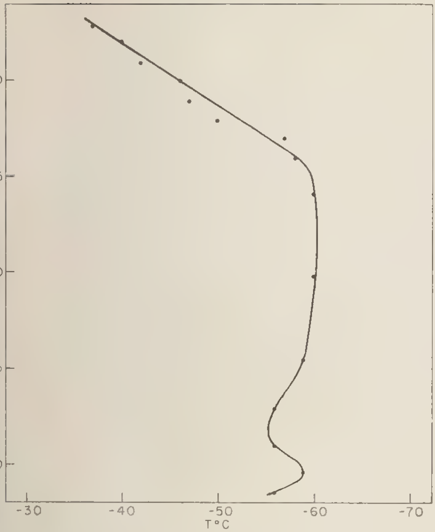


FIG. 10—Mean vertical temperature distribution with all observations from stations in Figure 1.

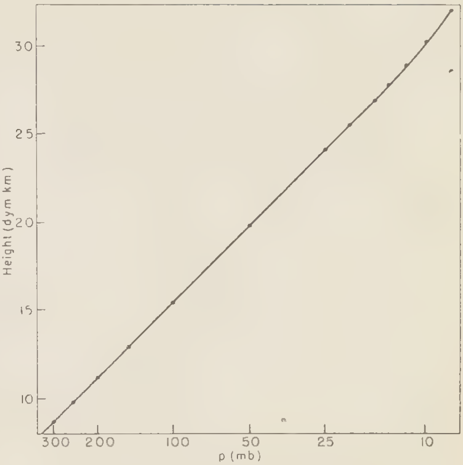


FIG. 11—Mean pressure-height curve with all observations from stations in Figure 1. Pressure plotted on logarithmic scale.

erally observed in jet streams, horizontal shears were about one-third larger to the left than to the right of the axis.

The c_n section (Fig. 8) is marked by cross-axis flow of 10 to 20 knots from lower to higher latitudes throughout. Since the averaging does not extend around a closed curve, this component does not necessarily represent the ageostrophic circulation about the axis. An attempt may be made to determine the latter by subtracting the geostrophic normal component c_{ng}, computed from the height difference at standard isobaric surfaces between 60°W and 170°W in the jet coordinate system, from the c_n field; the ageostrophic residual c_{na} is obtained (Fig. 9). As is found in other jet streams, this component is very weak compared with c_s. Figure 9 is therefore subject to large computational uncertainties; even if entirely correct, it can be taken as representative only for the period of calculation. Nevertheless, it is of interest that a complex pattern is found, with lateral entrainment toward the jet axis, rather than a simple transverse circulation from warm to cold air.

Mass and thermal field—In order to present these fields, a mean sounding was first computed from all available observations. Then deviations from the mean soundings were calculated and analyzed. In the average sounding the temperature was constant to 26 geopotential km, exactly the level of maximum wind (Fig. 10). Above this level the increase to the upper ozone layer began. Thus there was a pronounced increase in the stability of the atmosphere at the altitude of the jet core, comparable to the tropopause region associated with lower-altitude jet streams. There is a linear relation between log p and height up to 26 km; higher there was a change in slope comparable to that which is found in passing from troposphere to stratosphere (Fig. 11).

The cross section of height anomaly of isobaric surfaces (Fig. 12) with respect to the mean pressure-height curve of Figure 11 suggests a quasi-geostrophic c_s field. This is borne out by comparing the section of c_{sg}, the parallel geostrophic component (Fig. 13), with the observed wind, c_s (Fig. 7). Scatter diagrams of c_s and c_{sg} (Fig. 14) show an excellent correlation at 100 mb and 50 mb. The correlation is somewhat weaker at 25 mb, where, however, the scarcity of observations may be responsible for the result.

In Figure 7, the lateral shear above the core is much less than that below. This may be due merely to the fact that very few observations extended above 25 mb. Nevertheless, it is of

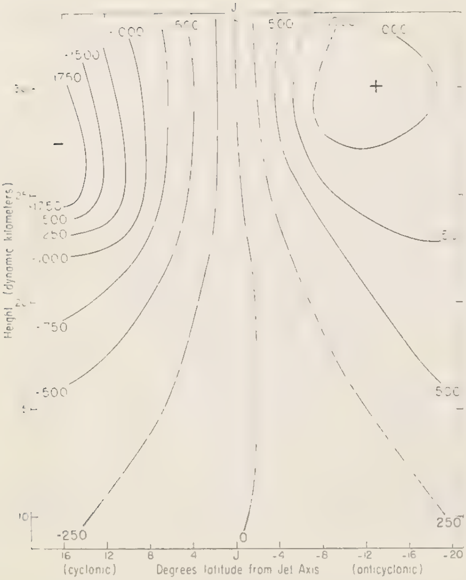


FIG. 12—Vertical cross section of altimeter correc- tion (meters).

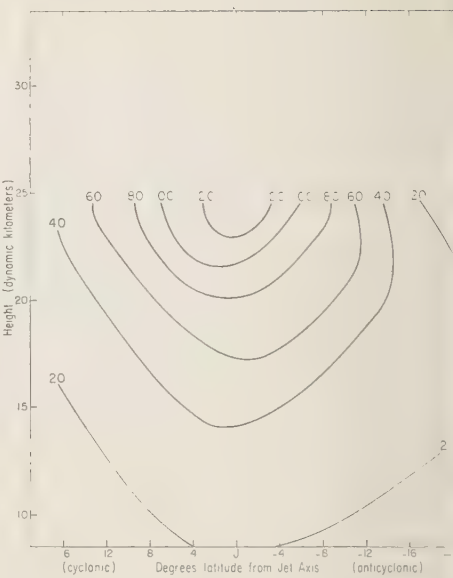


FIG. 13—Vertical cross section of geostrophic wi component parallel to jet axis c_{sg} (knots).

interest that the temperature-anomaly section (Fig. 15) supports this feature. Below the level of strongest wind, there was a deep layer, nearly isothermal, with a large temperature drop from right to left across the axis. Above the core, a reversal of the temperature field was only weakly indicated.

The field of potential temperature (Fig. 16) closely resembles that known from other jet streams.

Vorticity cross sections—The foregoing calculations can be used to determine the fields of absolute and potential vorticity. These have been measured on isentropic surfaces (Fig. 16) because the potential vorticity, when computed on these surfaces, is a conservative air-mass property except for non-adiabatic and frictional effects [*Ertel*, 1942]. In the stratosphere, these can be expected to modify the potential vorticity field only gradually.

The absolute vorticity was defined by

$$\zeta_a = f - \frac{\partial c_s}{\partial n} + \frac{u}{a} \tan \phi$$

where f is the Coriolis parameter, u the zonal wind component, a the radius of the earth, and ϕ latitude. The contribution to vorticity th arises from flow curvature was assumed cancel in the averaging from ridge to troug.

As may be expected from Figure 7, the gra ent of absolute vorticity was concentrated the jet axis (Fig. 17). A vorticity maximum w located just to the left of the jet core, thou it is weaker than is normally observed alc lower-altitude currents. The main difference tween these currents and the arctic jet stream the lack of a pronounced vorticity minim along the latter's right margin; rather, a c tinuous slow decrease in vorticity is found ward the equator.

Finally the potential vorticity was calcula (Fig. 18) using the definition

$$\zeta_p = \zeta_a]_\theta g \frac{\partial \theta}{\partial p}$$

where g is the acceleration of gravity and θ potential temperature. As is readily appar from Figures 16 and 17, this quantity cannot constant across the jet axis, since, just as in case of other jet streams, isentropic thickne decrease as the absolute vorticity increa Hence, one cannot regard the arctic jet str as being formed in connection with converge and divergence in an isentropic surface in a fluid with initially constant potential vorticity along the isentropes. Such formation would demand large isentropic thicknesses to the left and shallow thicknesses to the right of the axis. The strong gradient in potential vorticity across the jet axis must arise from the distribution of nonconservative influences, heat sources, and friction.

A complete explanation of this current must await observational and theoretical studies of these two effects.

Conclusion—The procedure employed in this paper is quantitative except for the initial loca-

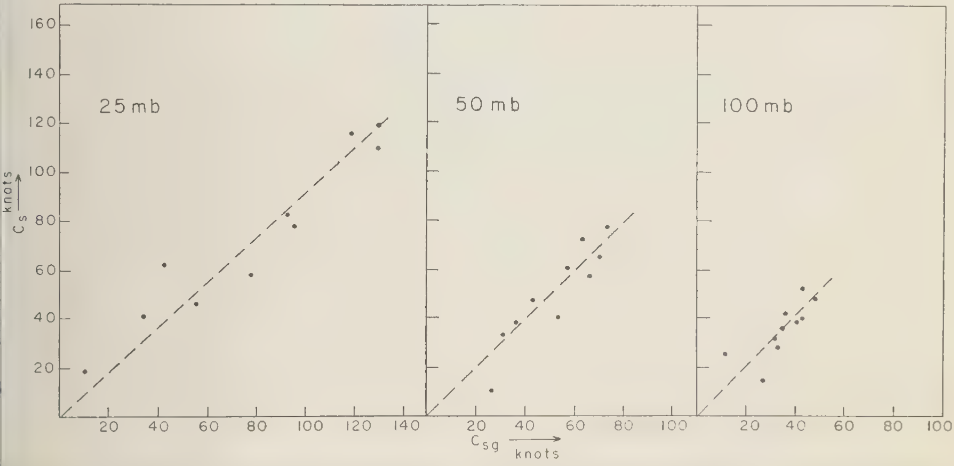


Fig. 14—Correlation between actual and geostrophic wind components parallel to jet axis at 100, 50, and 25 mb.

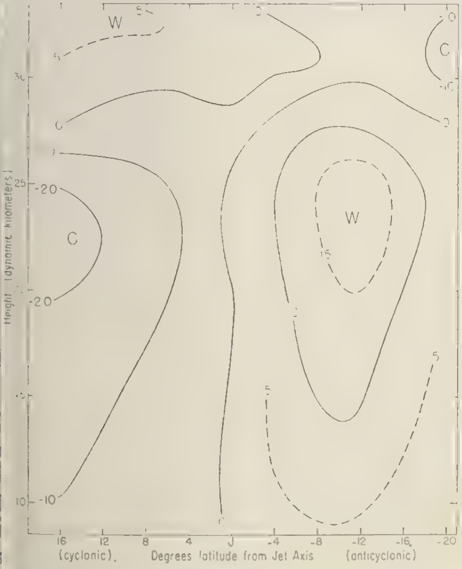


Fig. 15—Vertical cross section of temperature anomaly (°C) from sounding of Figure 10.

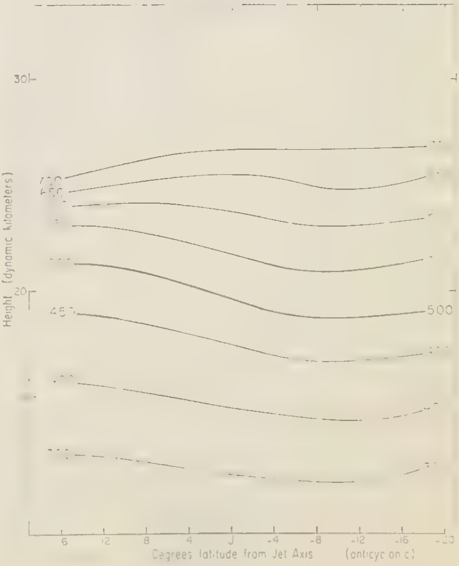


Fig. 16—Vertical cross section of potential temperature (°Kelvin).

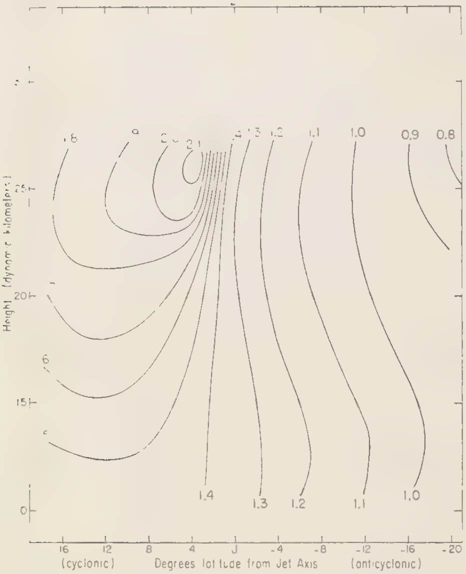


Fig. 17—Vertical cross section of absolute vorticity (10^{-4}sec^{-1}) computed along the isentropic surfaces of Figure 16 with formula given in text.

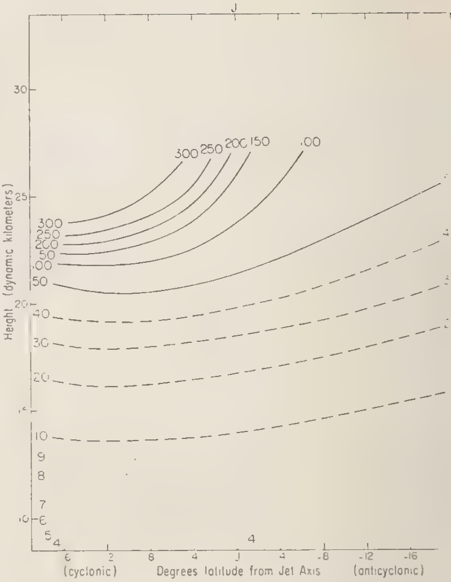


Fig. 18—Vertical cross section of potential vc ticity (10^{-2}cm sec^{-3} deg mb^{-1}) computed from Fi ures 16 and 17.

tion of the jet stream axes at 50 mb (Figs. 2-5). Therefore the cross sections obtained here are reproducible. Some of the results may be questionable because of some uncertainty about the reliability of the measuring instruments at high altitudes. This has not been discussed here. It must be noted, as it was remarked earlier, that the sections are for a specific four-day period. No claim is made that they are representative of all features of the polar-night jet stream at all times, particularly the altitude of the core. A climatological description of the current and its variations can only be obtained if the experiment carried out here is performed for a large number of periods in several seasons.

Acknowledgment—This work was carried out under a contract between the University of Chicago and the Office of Naval Research. The writer wishes to express his sincere appreciation to Herbert Riehl, who supervised the research.

References

Austin, J. M., and L. Krawitz, 50 millibar patterns and their relation to tropospheric changes, *J. Meteorol., 13,* 152-159, 1956.

Court, Arnold, Weather observations during 1940–41 at Little America III, *Proc. Am. Phil. Soc., 89,* 324–343, 1945.

Ertel, H., Ein neuer hydrodynamischer Wirb satz, *Meteorol. Z., 59,* 277–281, 1942.

Godson, W. L., and R. Lee, The arctic stra spheric jet stream during the winter 1955–56, *Meteorol., 14,* 126–135, 1957.

Godson, W. L., and R. Lee, High level fields wind and temperature over the Canadian Arct *Beitr. Phys. Atm., 31,* 40–68, 1958.

Heastie, H., *Average height of the standard i baric surfaces over the area from the North P to 55°N in January,* (M.R.P. No. 918) Meteor Research Comm., London, 17 pp., 1955.

Hess, S. L., Some new mean meridional cross s tions through the atmosphere, *J. Meteorol.,* 293–300, 1948.

Kochanski, A., Cross sections of the mean zo flow and temperature along 80°W, *J. Meteor 12,* 95–106, 1955.

Loewe, F., and V. Radok, A meridional aerolo cal cross section in the Southwest Pacific, *Meteorol., 7,* 58–65, 1950.

McIntyre, D. P., On the barocline structure the westerlies, *J. Meteorol., 12,* 201–210, 195

Palmén, E., Ober die Temperaturverteilung in Stratosphare und ihren Einfluss auf die Dynar des Welters, *Meteorol. Z., 51,* 17–23, 1934.

Teweles, S., Anomalous warming of the stra sphere over North America in early 1 *Monthly Weather Rev., 86,* 377–396, 1958.

(Manuscript received April 18, 1959; revised gust 12, 1959.)

The Use of Transosonde Data as an Aid to Analysis and Forecasting during the Winter of 1958-1959

J. K. Angell

U. S. Weather Bureau
Washington, D. C.

Abstract—During the winter of 1958–1959, transosonde positions and transosonde-derived winds were plotted routinely on the 250-mb chart by the National Weather Analysis Center (NWAC) at Suitland, Maryland. A statistical assessment is presented of the quantity and quality of transosonde data so plotted. The usefulness of transosonde data in specific instances is pointed out.

Introduction—Between September 1957 and April 1959, the United States Navy flew constant-level balloons (transosondes) from the Naval Air Station at Iwakuni, Japan, on an operational basis. Summaries of these flights have been given by *Angell* [1959a,b]. Between November 1958 and April 1959, the transosonde positions and transosonde-derived wind data were plotted on 250-mb Northern Hemisphere maps issued by the National Weather Analysis Center (NWAC) at Suitland, Maryland. The purpose of this paper is to show the usefulness of these transosonde data in filling the meteorological void over the Pacific Ocean and to indicate the cost involved in obtaining meterological data by this technique.

The transosonde system instrumentation and procedures—The transosonde balloon in use between 1957 and 1959 was 40 ft in diameter and had a volume of about 34,000 cu ft (Fig. 1). It was a nonexpansible balloon made of 1.5 mil polyethylene laminate. Since the balloon was nonexpansible, its natural floating level was determined by the balance between bouyancy force (a function of environmental temperature) and the total weight of equipment carried. Owing to the continual seepage of helium through the skin, the balloon could be maintained at a constant elevation only by means of a ballast system activated by a barograph. Of the 600 to 700 pounds carried to 300 or 250 mb by this balloon, about 400 pounds consisted of ballast. The tendency for the natural floating level of the balloon to ascend as the balloon dropped ballast was counteracted in 1958 by the addition of a fan which mixed air with the helium in the balloon and thus reduced the lift. With the addition of this refinement, the transosondes seldom deviated more than 1000 ft from their prescribed flight altitude.

For the purpose of telemetering information, each transosonde was equipped with a 50-watt transmitter which operated for 15 minutes every 2 hours at alternate frequencies of 6, 13, and 19 Mc/s. The transosonde position was

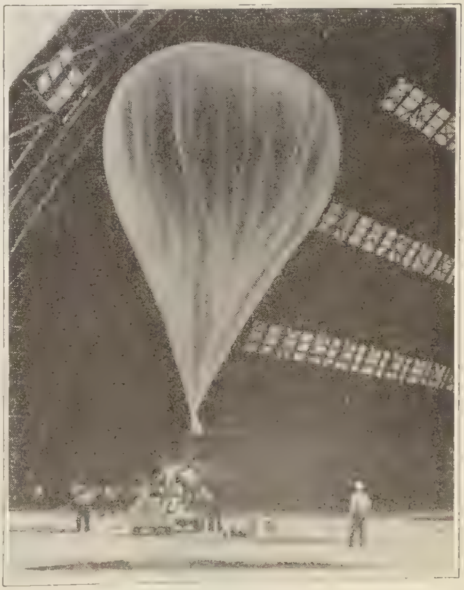
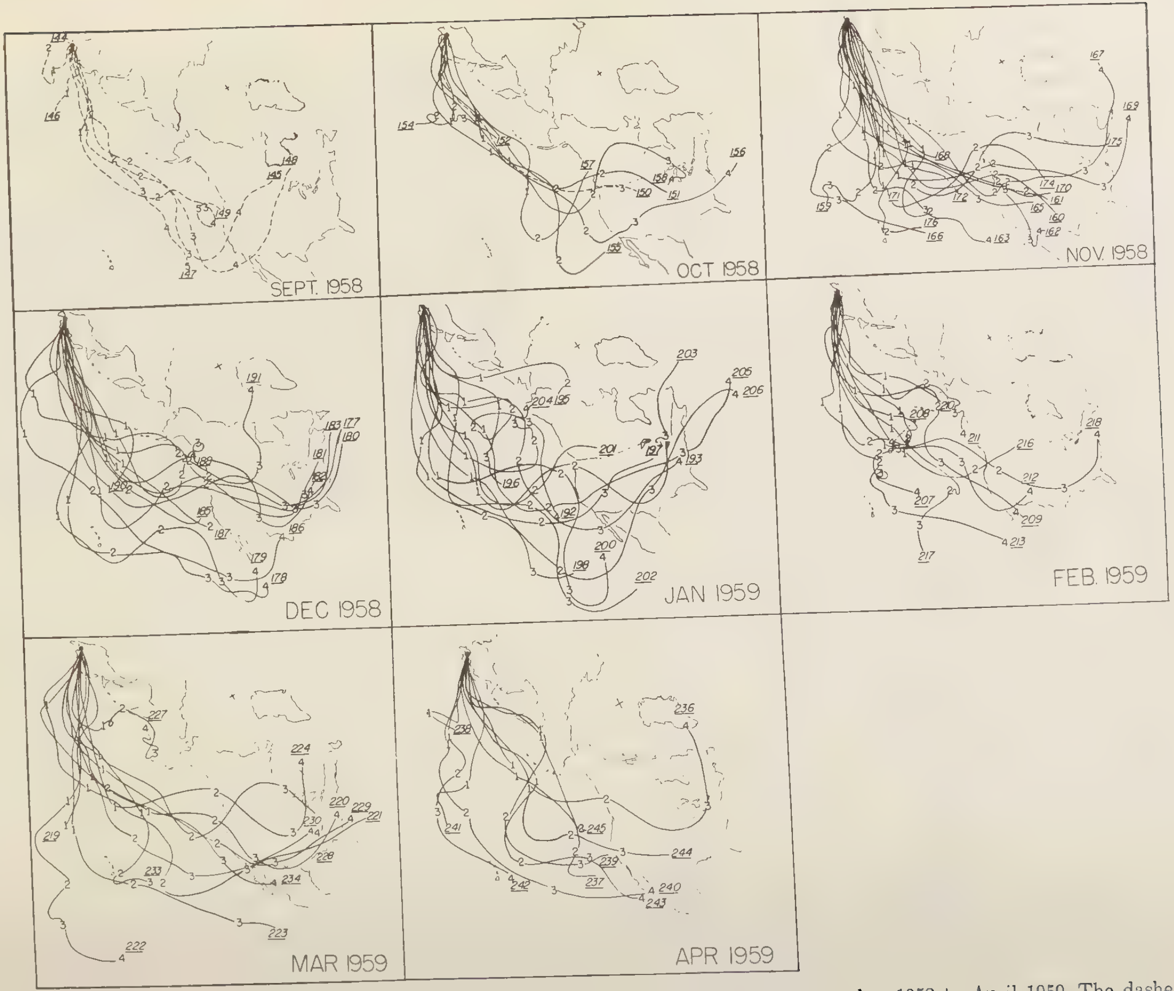


Fig. 1—Preparations for launching a transosonde balloon.



1958 to April 1959. The dashed

letermined by means of radio-direction-finding (RDF) bearings on these signals. This positioning was carried out by 20 RDF stations of the Federal Communications Commission (FCC) located within the United States (including Alaska and Hawaii) and by the two U. S. Navy RDF networks in the Pacific Ocean area. The bearings taken by individual RDF stations were analyzed in Washington, D. C., and Pearl Harbor, Hawaii, and a most probable transosonde position was determined to the nearest).1° of latitude and longitude. In addition, a rating of accuracy was given each fix based upon the area of intercept of the bearings. The wind vectors were derived from the RDF positions on the basis of distance and direction traveled by the transosonde as a function of time. For further discussion of the transosonde system see a report by *Anderson and others* [1955].

Transosonde flights during 1958–1959—Figure 2 shows the transosonde flights of one day's duration or more during 1958–1959. The dashed rajectories during September represent flights at 300 mb (30,000 ft) and the solid trajectories represent flights at 250 mb (34,000 ft). The light altitude of the transosondes was changed at the beginning of October in accord with a imilar change in the pressure surface being analyzed by NWAC. Of the 100 flights launched during 1958–1959, 88 were tracked for one day or more. However, even though transosonde lights of 7 days' duration were quite feasible, ill the flights made after September 1958 were imited to 4 days' duration in order to avoid nterference with the activities of commercial et aircraft over the Atlantic Ocean. Furthermore, the transosondes were not released from apan if there was much chance that they would pass over the Asiatic mainland. This combination of political and hazard-to-aircraft considerations severely restricted the transosonde potential.

The transosonde plotting model—As mentioned above, the operational transosonde flights rom Japan were positioned at 2-hour intervals y means of FCC and U. S. Navy RDF stations. The positions obtained, and the winds erived therefrom, were then placed on meteorological teletypewriter circuits in Pearl Harbor, Iawaii, and Norfolk, Virginia, for transmission o analysis centers where the data could be used operationally. However, owing to the difficulty of obtaining accurate transosonde positions over the vast reaches of the Pacific Ocean, it soon became apparent that the transmission of smoothed transosonde positions would be desirable. These smoothed positions were obtained by averaging the latitudes and longitudes of three successive transosonde positions 2 hours apart. Moreover, in order to obtain sufficiently reliable wind data from the transosonde positions, these winds were evaluated (at 2-hour intervals) from the distance and direction between smoothed positions 6 hours apart.

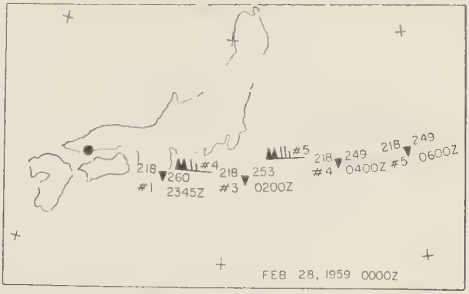
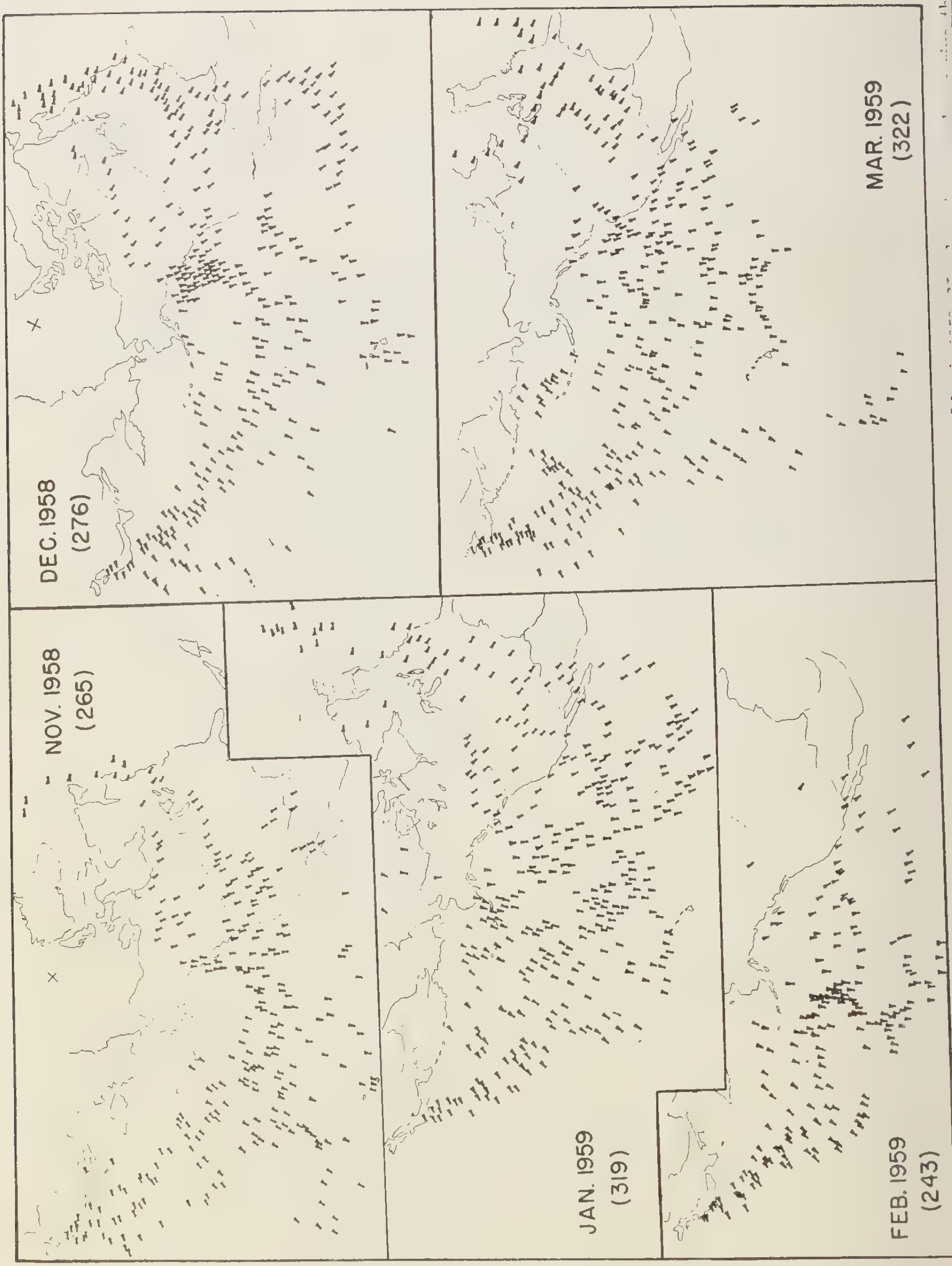


Fig. 3—The transosonde plotting model illustrated by reference to the NWAC plot for a portion of transosonde flight 218 in February 1959.

Figure 3 shows how the transosonde data transmitted by teletypewriter were plotted on the NWAC 250-mb 0000Z map of February 28, 1959. The smoothed transosonde positions are represented by dels, with the flight number plotted to the upper left of the del, the message number to the lower left, the pressure at which the transosonde was flying to the upper right, and the time of the position to the lower right. The transosonde-derived wind vector was plotted midway between the two smoothed positions from which it was evaluated, with the message number, and sometimes the flight number, attached. In general, smoothed transosonde positions 6 hours later than map time and transosonde-derived winds 3 hours later than map time were available to plot on the NWAC 250-mb (Fig. 3). Thus the transosonde data were fairly well centered (timewise) on the NWAC 250-mb maps.

Operational use of transosonde data during 1958–1959—Figure 4 shows the transosonde positions plotted on NWAC 250-mb maps between


DEC. 1958
(276)
NOV. 1958
(265)
JAN. 1959
(319)
FEB. 1959
(243)
MAR. 1959
(322)

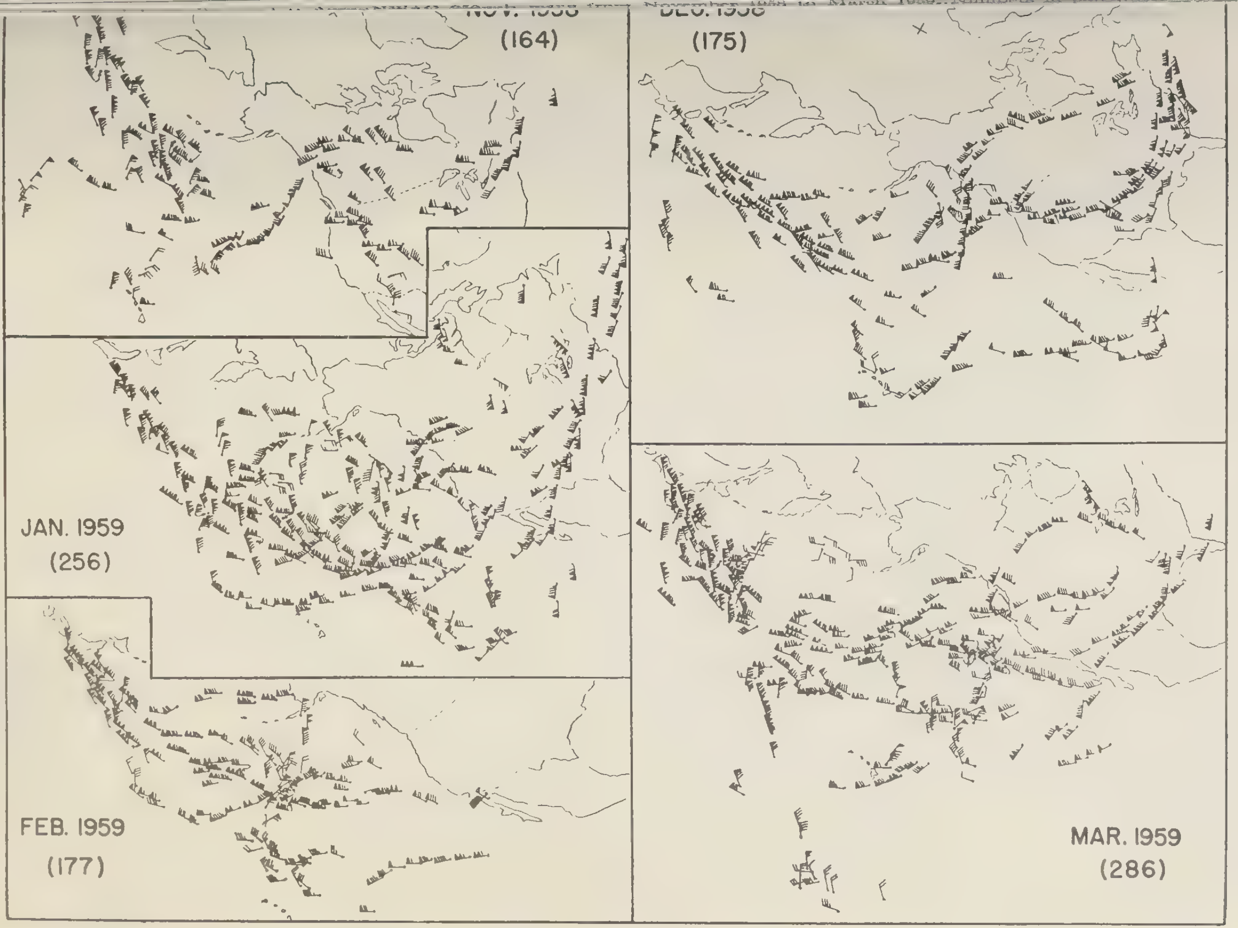


FIG. 5—Transosonde-derived winds plotted on NWAC 250-mb maps from November 1958 to March 1959. Numbers in parentheses give the total plottings for each month.

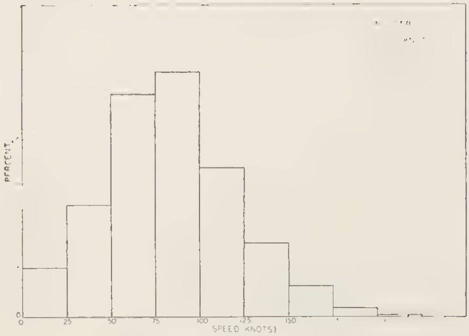


Fig. 6—Distribution of transosonde-derived wind speeds plotted on NWAC 250-mb maps between November 1958 and March 1959.

November 1958 and March 1959. The numbers in parentheses give the number of transosonde positions plotted during each month. In the 5-month period, a total of 1425 such (smoothed) positions were plotted. Figure 4 shows that during most of the months the positions were well scattered over the Pacific Ocean with no large blank spots. Figure 5 shows the 6-hour-average transosonde-derived winds plotted on NWAC 250-mb maps between November 1958 and March 1959. In the 5-month period a total of 1058 transosonde-derived winds were plotted, yielding wind data over vast reaches of the Pacific Ocean where previously such data had been unavailable. As a matter of interest, Figure 6 shows the magnitudes of the wind speeds plotted in Figure 5. The mean 6-hour-average wind speed was 81 knots and the mode was 75 to 100 knots. About 5 per cent of the wind speeds exceeded 150 knots. This percentage is not higher because the use of a 6-hour averaging procedure severely smooths out the peak wind speeds.

For the purpose of emphasizing the usefulness of transosonde data in the analysis of individual synoptic maps, let us first take a case where three transosondes were aloft over the Pacific Ocean at the same time. Figure 7 represents a copy of the NWAC 250-mb 0000Z map of February 28, 1959. Transosonde flights 216, 217, and 218 were released from Iwakuni, Japan, within 24 hours of each other, and in this figure they are shown spread out across the Pacific Ocean in a condition of nearly zonal flow. The transosonde data coverage would be even more stri ing except for the fact that flight 217 could n be positioned until it was considerably east Japan, as shown by the transosonde positi indicated in message 2. In general, the m analyst has used the transosonde data to go advantage in making the 250-mb analysis. No that on flight 216 the transosonde-derived wi speed decreased from 175 knots to 70 knots 8 hours. Such a deceleration would be associat (from the equation of motion) with a negati angle between wind and geostrophic wind 18° (flow toward high pressure) so that, wheth by design or accident, the analyst has proper suggested that the transosonde is moving acro contours toward high pressure. What is n explained from the contour pattern, as draw is why the transosonde is experiencing such strong deceleration to begin with, since t contour gradient is not shown as weakeni downstream.

In March 1959, the transosonde messag were first transmitted on a scheduled basis the teletypewriter circuits. This resulted considerable improvement in the number messages reaching NWAC. As an illustration the quantity and quality of data obtained aft this changeover and the resultant usefulne of even one transosonde trajectory, a porti of the trajectory of flight 239 plotted on se ments of four NWAC 250-mb maps is sho in Figure 8. In this case the trajectory serves delineate the trough position in the Gulf Alaska and also gives an estimate of the wi speed around this trough. The succession transosonde-derived wind speeds plotted these four maps indicates the regularity of accelerations and decelerations of the wind be expected at such altitudes. As the transonde passes close to weather ship *N* (00(map of April 9) it is seen that the 85-knot w derived from the transosonde displacement agr well with the 80-knot wind reported by weather ship. The analyst has used the tran sonde data to good advantage on these f maps, except perhaps on the 1200Z map April 9, 1959, where the northward mover of the transosonde casts doubt on the loca of the 250-mb ridge so far to the west.

The percentage of transosonde-derived c which appeared on the NWAC 250-mb m

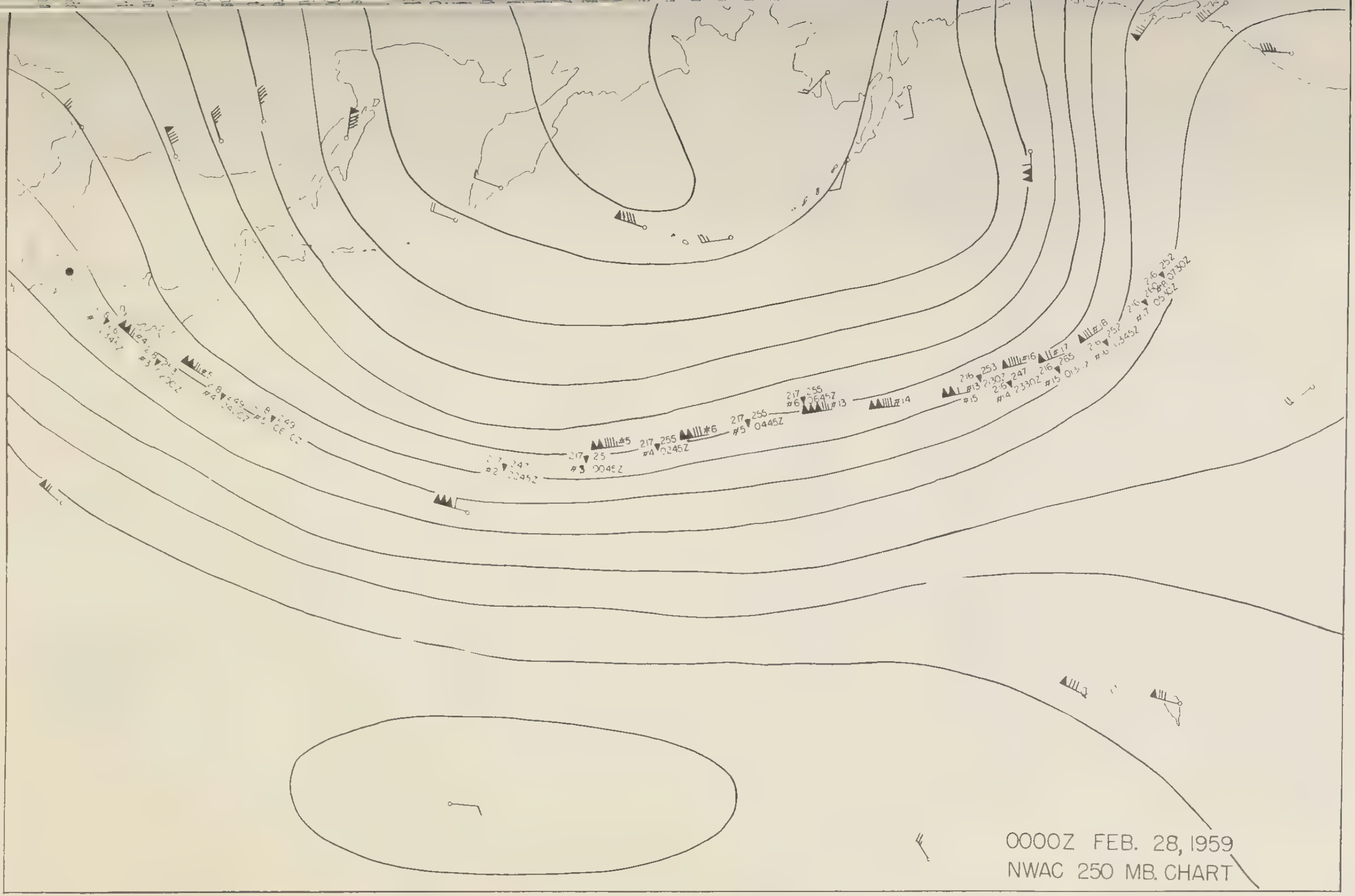


FIG. 7—Copy of NWAC 250-mb map with plots for transosonde flights 216, 217, and 218 in February 1959. Contours drawn at 400-ft intervals.

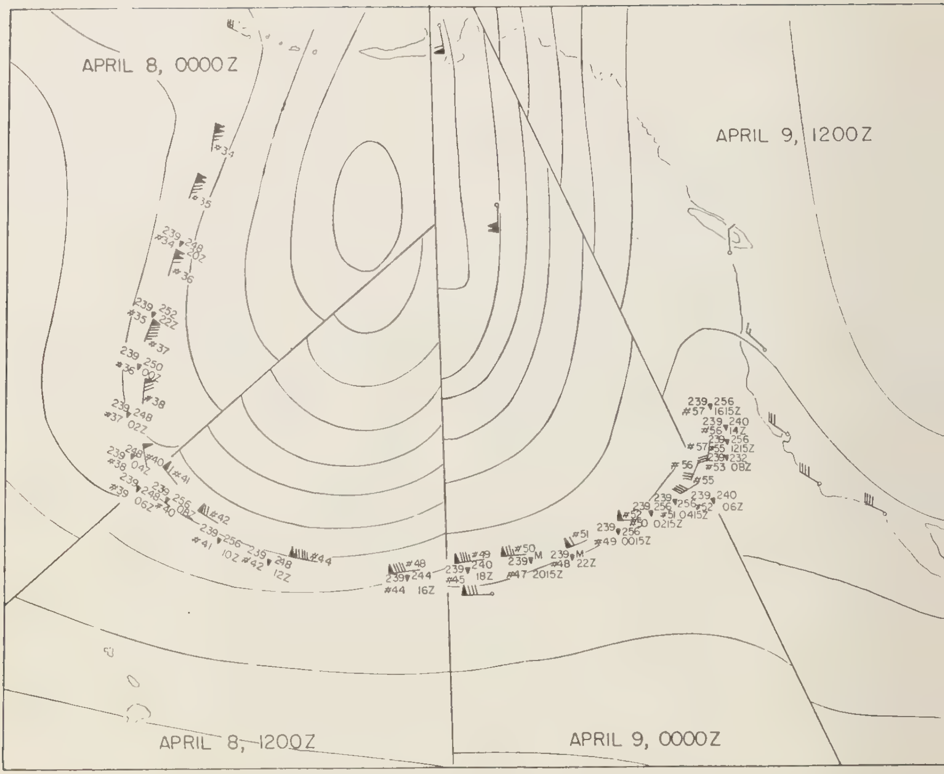


Fig. 8—Portion of the trajectory of transosonde flight 239 plotted on segments of four NWAC 250-m maps in April 1959. Contours drawn at 400-ft intervals.

increased through the winter of 1958–1959. Figure 9 shows the percentage of possible transosonde positions (solid line) and transosonde-derived winds (dashed line) plotted on NWAC maps between November 1958 and March 1959. The word 'possible' refers to the total number of 2-hour positions and winds which could have been plotted had the tracking, communications, and plotting procedures functioned perfectly. The percentage of possible positions plotted increased gradually from 50 per cent in November to 60 per cent in March. The percentage of possible transosonde-derived winds plotted increased more abruptly from 30 per cent in November to 60 per cent in March. Apparently the analysts and forecasters found the transosonde-derived winds useful during the winter and became more insistent upon their being plotted on the NWAC 250-mb maps.

Cost estimates of transosonde-derived data- The foregoing shows that useful data can l obtained over the Pacific Ocean by means the transosonde system. Whether such data a obtained in the future depends upon the pri weather services are willing to pay. As of M 1959, the U. S. Navy has decided against co tinuation of operational transosonde fligh largely on the basis of expense. In order indicate the cost of obtaining data by t transosonde system, estimates are present (Table 1) of the cost of each transosonc derived wind plotted on NWAC 250-mb ma during the winter of 1958–1959 and the cost various conditions had been realized. With t various alternatives, the cost per wind detern nation varies from $300 to $75. There will be attempt here to compare these costs with t cost of obtaining wind data from weather shi

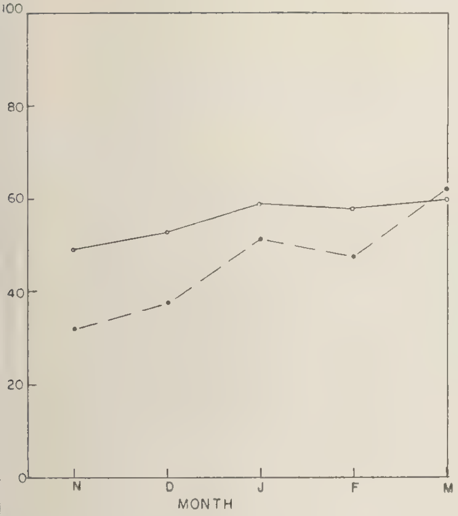


FIG. 9—Percentage of possible transosonde posi-.ons (solid line) and transosonde-derived winds dashed line) plotted on NWAC 250-mb maps ·om November 1958 to March 1959.

nce the weather ships obtain winds, tempera- ıres, and pressures at various levels, and also ·rve as rescue units in cases of emergency. It ıould be mentioned, however, that if a rela- vely cheap positioning system for the transo- ɔndes could be found, the cost per wind de- ·rmination indicated in Table 1 would be ·duced by one-third, and the economics of the ·ansosonde system would appear more favor- ble.

Conclusion—During the summer of 1959, su- ·rpressure balloons were tested. These balloons ın maintain constant-level flight without need ɔr a ballast system, and consequently the equip- ıent can be made cheaper and much lighter ıan that heretofore in use. These superpressure ılloons represent almost no hazard to aircraft ıd are probably capable of circumnavigating the hemisphere. Whether such flights take place in the future is dependent upon international and financial considerations. It is believed that, on the basis of the usefulness of the transosonde data obtained during 1958–1959, a very good case can be made for the furtherance of the transosonde concept both for operational and research purposes.

TABLE 1—*Cost estimates of a single transosonde-derived wind vector plotted on NWAC 250-mb maps*

Actual cost of each transosonde-derived wind plotted during the winter of 1958-1959	$300
Estimated cost of a single transosonde-derived wind if all 1958-1959 transosonde flights were of 4 days' duration.	$225
Estimated cost of a single transosonde-derived wind if all 1958-1959 transosonde flights were of 4 days' duration and the tracking, communications, and plotting functions were performed perfectly.	$112
Estimated cost of a single transosonde-derived wind if all 1958-1959 transosonde flights were of 8 days' duration and the tracking, communications, and plotting functions were performed perfectly.	$75

REFERENCES

ANDERSON, A. D., H. J. MASTENBROOK, AND H. D. CUBBAGE, *The transosonde—a new meteorological data-gathering system,* Rept. 4649, Naval Research Laboratory, Washington, D. C., 1955.

ANGELL, J. K., *A summary of Navy-sponsored 300-mb constant level balloon (transosonde) flights from Japan in 1957–'58,* unpublished manuscript, U. S. Department of Commerce, Weather Bureau, Washington, D. C., 1959a.

ANGELL, J. K., *A summary of Navy-sponsored 250-mb constant level balloon (transosonde) flights from Japan in 1958–'59,* unpublished manuscript, U. S. Department of Commerce, Weather Bureau, Washington, D. C., 1959b.

(Manuscript received July 13, 1959; presented at the fortieth Annual Meeting, Washington, D. C., May 6, 1959.)

Application of Meteorological Rocket Systems

WILLIS L. WEBB AND KENNETH R. JENKINS

U. S. Army Signal Missile Support Agency
White Sands Missile Range, New Mexico

Abstract—A series of test rocket firings has been conducted during the past 18 months to establish the operational feasibility of numerous rocket systems for meteorological observations. As might be expected, the most desirable systems from the point of view of instrumentation are generally not the most desirable from that of the rocket-firing problem. It has been demonstrated, however, that a reasonable observation schedule can be accomplished by the judicious application of currently available rockets and sensors. The most variable of high-atmosphere meteorological parameters is the flow. Chaff was used initially for rocket wind measurements because it could be expected to provide a suitable indication of the wind in the atmosphere above balloon sounding levels. It is easy to package and deploy. Most of the available high-atmosphere wind data have been obtained through use of a chaff sensor, and it is still most applicable for point measurements and at very high altitudes.

The need for a more coherent sensor and a vehicle capable of transporting a telemetry system to provide for the measurement of other parameters has resulted in the development of a parachute system. Although an altitude range problem will always be encountered, it is possible to obtain data from approximately 200,000 ft. to the surface through the application of a single parachute and balloon combination. Launch and flight characteristics of the tested rockets are presented for use in applying this new observational technique. Careful adherence to the design and operational restrictions indicated by these data will result in savings in the effort required for development of the various desirable measuring techniques. Experience to date indicates that it is possible, with available equipment and a reasonable expenditure of effort, to obtain profiles of several meteorological parameters from the surface to altitudes of the order of 200,000 ft.

Introduction—The lower reaches of the atmosphere have received a great deal of attention during the past 30 years as a result of exploration by balloon techniques. A large amount of data has been accumulated between the surface and 50,000 ft by these methods, but the data are usually more sparse at higher altitudes and are generally not considered satisfactory above 75,000 ft [*Mervill,* 1949]. It is possible to increase the maximum altitude obtained by balloons, but a significant amount of data will probably not be obtained above 100,000 ft through balloon systems.

The application of satellite vehicles to meteorological observation problems is opening a new era for meteorologists. Although the earth satellite performs a function of which no other system is capable, it is limited in that it cannot operate in the regions of the atmosphere below 100 miles. The interim region of the atmosphere is, therefore, largely unprobed by direct observation. Although observations from below or from above may produce significant information it is not likely that the necessity of direct observation can be avoided. The only available system for systematically probing the region from 20 miles to 100 miles is the meteorological rocket [*Stroud,* 1958; and *Spencer,* 1958]. A rocket technique is neither simple nor easy, but in view of the complete lack of other methods of obtaining data, the details of such an operation are presented for the edification of the profession.

The U. S. Army Signal Missile Support Agency, White Sands Missile Range, New Mexico, has initiated observational studies of the atmosphere by rocket systems to complement its activities in support of the missile program [*Jenkins and Webb,* 1958]. The effects of the atmosphere on the flight of a missile cover a wide range of physical phenomena. The ballistic effects of drag and wind must be considered, as well as the propagation effects involved in the transmission of various forms of energy that the missile communicates to the atmosphere, either

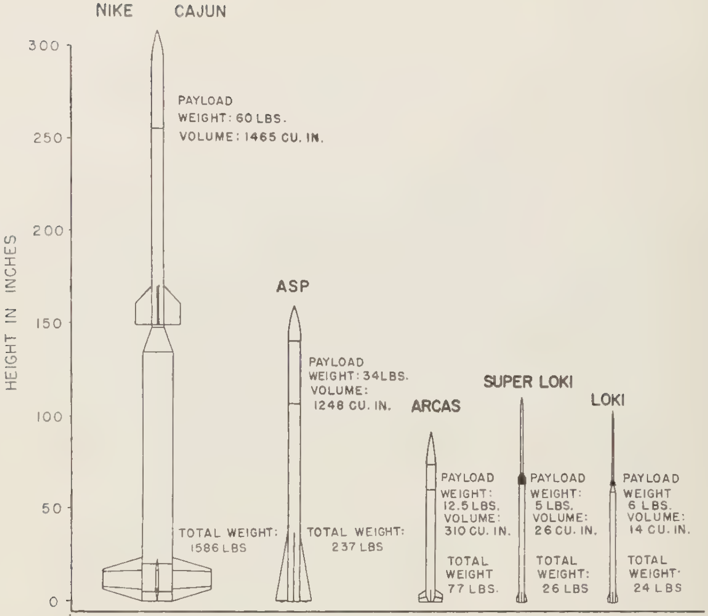


Fig. 1—Configurations and performance characteristics of tested meteorological rockets.

as desired or as inadvertent radiation. A meteorological rocket launching installation has been set up, and a total of 120 rounds has been fired in support of various experiments requiring information about the state of the upper atmosphere. The following data are presented to show the meteorologist the scope of problems encountered and to indicate the present state of the meteorological rocket art.

General performance characteristics of tested meteorological rockets—A large variety of rockets has been utilized in atmospheric studies in the past 10 years [*Haig and Lally*, 1958]. This discussion will be devoted to the smaller types useful in a synoptic system. They range from the small Loki rocket to the Nike Cajun configuration. The choice of a rocket for use in a particular case rests largely on the payload requirement and the peak altitude desired. The small rockets are more easily handled and are desirable wherever they can meet the experimental requirements.

As is indicated in the configuration and performance data (Figs. 1 and 2), the Loki Phase I can be fired with a total rocket weight of only 24 lb to carry a 2-lb payload to approximately 140,000 ft at White Sands Missile Range. Th
Loki Phase IIA can deliver the same payload t
280,000 ft. The stress on the Loki Phase II
system is such that an alternative vehicle ha
been developed. The system consists of the Lok
Phase I booster and an enlarged dart whic
carries a 2-lb Naka motor. This system is expected to provide adequate altitudes withou
the excessive speeds that result in gross aerodynamic heating. The Arcas rocket was develope
by the U. S. Navy, Office of Naval Research, t
meet the need for an economical, easily handle
meteorological rocket with an adequate payloa
[*Webb, Jenkins, and Clark*, 1959]. As can b
seen from the performance characteristics, th
Arcas delivers a 12½-lb payload to peak alt
tude with a total lift-off weight of 77 lb.
can also be classed as an easily handled vehicl

If additional payload in weight or volume
required, the Asp offers a reasonable solutio
Its total weight of 237 lb makes the launchi
operation more difficult. Although the As
payload capability is advantageous, its perfor
ance and cost rule out its application to synopt
observational programs.

The final meteorological rocket system d

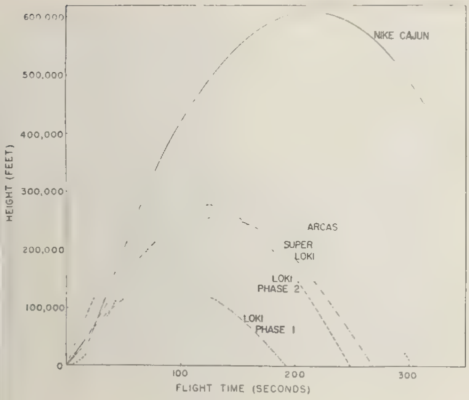


Fig. 2—Idealized trajectories of tested meteorological rockets.

cussed is the Nike Cajun configuration. This vehicle is composed of a Cajun motor and a Nike Ajax booster, provided with suitable fin modifications. The booster is used to propel the second-stage Cajun rocket to an altitude of approximately 50,000 ft, where the sustainer motor can operate efficiently. A payload of 60 lb contained in 465 cu in can be lifted to above 100 miles with the Nike Cajun. The rocket's total weight when ready to leave the rail is 1566 lb. The application of the Nike Cajun rocket is limited by the cost and the magnitude of the problems associated with preparation and firing.

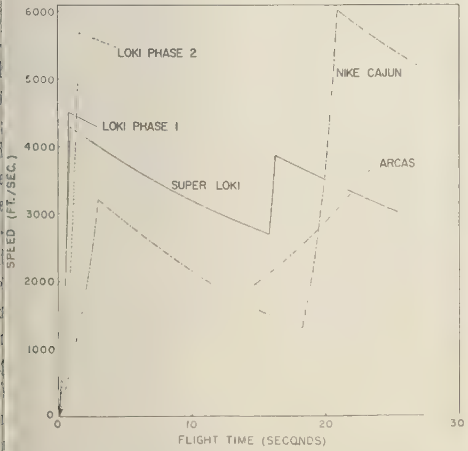


Fig. 3—Velocity distribution for selected meteorological rockets.

Idealized trajectories are presented, Figure 2, for the Loki Phase I, Loki Phase IIA, Super Loki, Arcas, and Nike Cajun. The Loki Phase I reaches a peak height of approximately 140,000 ft when fired from the 4000-ft altitude of the White Sands Missile Range. The Loki Phase IIA coasts to approximately 280,000 ft, which is similar to the planned performance of the Arcas. The Super Loki is designed for a peak altitude above 200,000 ft, and the Nike Cajun reaches well above 500,000 ft. As can be observed from these graphs, the time of sensor exposure on board the rocket is limited at any altitude with any of the vehicles. The maximum period of observation is obtained near peak and thus provides opportunity for more extended observations at that altitude.

Figure 3 indicates the velocity distribution expected during the burning phase, or phases, of the several vehicles. The Loki series is seen to experience large accelerations during the initial phase of the flight, and thus the structural stresses and thermal inputs are at a maximum. Accelerations are in excess of 200*g*, requiring relatively rugged instrumentation. The Nike

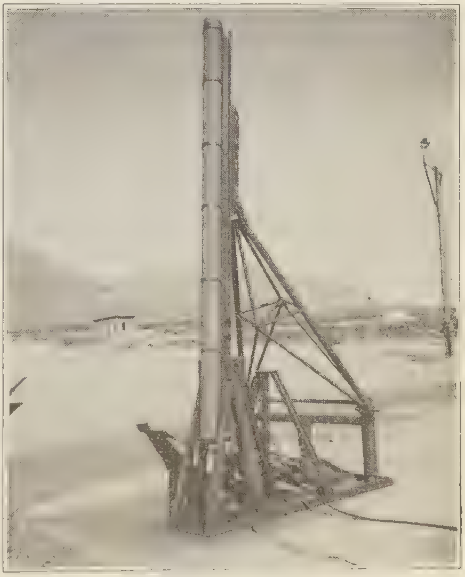


Fig. 4—The Loki launcher in firing position.

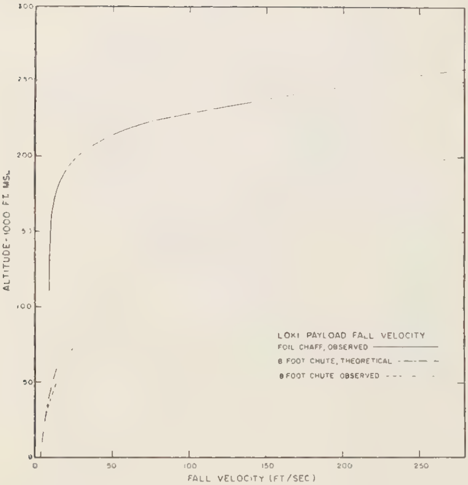


Fig. 5—Chaff and parachute wind sensor fall velocities.

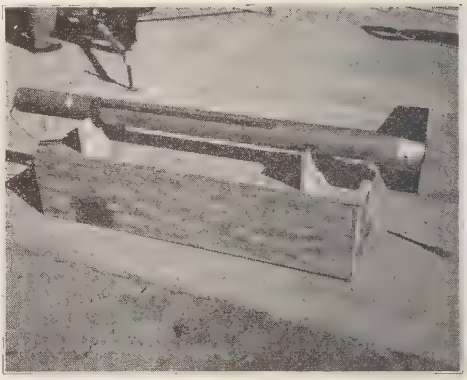


Fig. 6—Arcas meteorological rocket being prepared for beacon performance test.

Cajun is a short-burning rocket with accelerations of the order of 50g during boost phase and second-stage burning. The provision for an extended coast phase in the Super Loki and Nike Cajun permits second-stage ignition at a high altitude, taking advantage of the low drag at that stage of propulsion. The Arcas, on the other hand, uses a slow-burning motor which provides low acceleration over a long period of time to achieve a reasonable burnout velocity at an altitude where the drag is relatively low. Reliable instrumentation in this rocket is relatively easy to achieve, owing to the low accelerations involved.

The Loki system—The Loki meteorological rocket, a modification of the Loki tactical missile, is launched from the extruded-rail tubular launcher shown in Figure 4. Burning time for the Loki Phase I is 8/10 sec, at which time the dart is traveling in excess of 4400 ft sec. The Loki has two major components. The first stage contains solid propellant and provides the total thrust for the rocket flight. The initial acceleration causes the second-stage dart to turn in a J slot and free the locking pin. Separation is achieved at burnout through the force of a spring-loaded piston and differential drag. The initial acceleration causes a weight in the tip of the dart to shear a safety pin, igniting a percussion cap, and initiating a pyrotechnic train which is set for the desired time of payload expulsion The dart coasts to peak altitude and the winc sensor is ejected. The sensors that have beer tested extensively consist of an 8-ft mylar parachute, metalized for radar tracking, for use with the Loki Phase I vehicle, and radar reflective chaff [*Thaler and Masterson,* 1956; anc *Vaughn,* 1957]. The parachute has provided ai excellent point source in the region from 140,00 ft down to 75,000 ft. Chaff has been used extensively in the Loki Phase IIA to obtain win measurements in the altitude range from 280,00 ft down to 140,000 ft. As can be observed i Figure 5, the fall rate of a sensor which is acceptable in one of these ranges will generall not be acceptable in the other. The parachut represents an excellent tracking target for th radar and provides a point source throughou its descent. Comparisons of the wind measurements with radiosonde values in balloon-attaine levels indicate that the parachute provides reasonable means for evaluating the winds. Conversely, the chaff load disperses with time afte ejection and the wind determination becom more difficult [*Anderson and Hoehne,* 195 *Anderson,* 1957; *aufm Kampe,* 1957; *Batta* 1958; *Cline,* 1957].

In addition, the strong winds frequently encountered aloft make it difficult to track o sensor through a 100,000-ft stratum. Conside able research will be necessary to achieve suitable sensor for the entire wind profile.

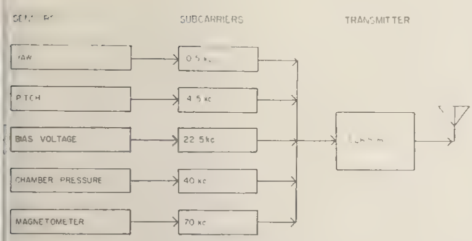


Fig. 7—Telemetry instrumentation for an Arcas performance test.

The Arcas system—The Arcas rocket was given initial flight testing at White Sands Missile Range, New Mexico, during the winter of 1958–1959. A series of 5 rounds was fired to establish the feasibility of the launching system and to determine the aerodynamic stability of the rocket. A view of the rocket during beacon checkout is shown in Figure 6. The initial firing incurred a structural failure which caused deviation from the planned trajectory at 15,000 ft. The nose cone was equipped with a DPN-43 radar beacon which failed at 15 sec, but skin tracking of the rocket by radar was possible. The second Arcas rocket was instrumented with a telemetry system in an attempt to establish the cause of the initial failure. The rocket was equipped with a 226.5-Mc/s 2-watt transmitter which was modulated by five subcarrier oscillators (Fig. 7). The subcarrier oscillators were in turn controlled by sensors which measured the pitch, yaw, combustion-chamber pressure, and a single component of the earth's magnetic field. The pitch and yaw were desired for evaluating aerodynamic stresses on the vehicle; the chamber pressure was needed to determine the thrust developed; and the magnetometer was included to measure the roll rate of the vehicle and to evaluate the attitude of the missile near peak where the parachute would be ejected. Figure 8 indicates the data obtained from the telemetry system during the burning phase. As can be observed, the combustion-chamber pressure behaved somewhat sporadically during the initial phases, operated smoothly through most of the burning phase, and was slightly unstable at burnout. The angle between the flow about the missile and the missile axis shows large deviations during the early phases which were probably due to the relatively small aerodynamic loads on the sensor at these speeds. The large excursions observed as the rocket approached the speed of sound were verified through reduction of ballistic camera data obtained in support of the firing. The large accelerations involved in the roll rate of the rocket as it became supersonic are not clearly understood.

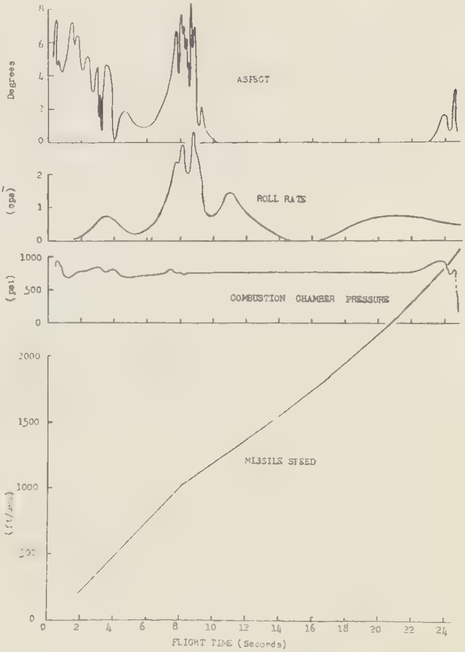


Fig. 8—Performance characteristics of an Arcas test round.

The third Arcas test round was fired, with an AMT-4 radiosonde transmitter as payload, to check the possibility that the extra drag resulting from the telemetry antennas and the performance sensors were the causes for the low peak latitude of 73,000 ft obtained with round 2. As round 3 reached a peak altitude of 93,000 ft, it was assumed that a further reduction in the drag was required. The roll rate of the rocket is presented in Figure 9. These data were obtained by recording the signal strength of the AMT-4 antenna located in the side of the nose cone. The signal strength was modulated as the rocket rolled and the antenna was turned toward and away from the receiver. The first few seconds of flight were unrecorded as a result of acquisition problems resulting from the GMD-1's failure to track in automatic position. The GMD-1 was 6 miles from the launch point. After manual acquisition at approximately 15 sec, automatic tracking was maintained throughout the flight.

The Arcas Launcher consists of an 11-ft tube which has a larger concentric chamber about its base. The rocket is mounted on a split piston and is held in alignment with the axis of the launcher by means of rigid foam blocks. This equipment falls away as the rocket leaves the launcher. Initially the exhaust gases are allowed to by-pass the piston to reduce lift-off accelerations. The pressure then builds up and assists in obtaining the desired exit velocity.

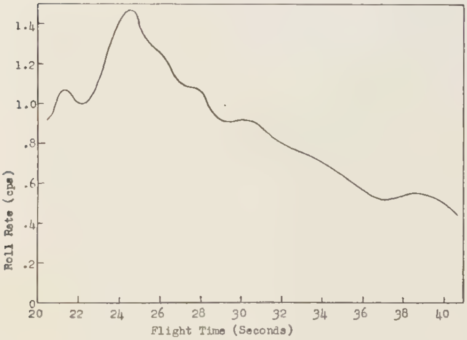


FIG. 9—Roll rate of the Arcas during test launching, obtained by using AMT-4 transmitter and GMD-1 ground equipment.

A redesign of the nose-cone configuration was incorporated in subsequent rockets, and the fourth round was fired with a dummy load to an altitude of 178,000 ft. The fifth flight, which achieved an altitude of 171,000 ft, included an AMT-4 radiosonde transmitter which operated only during the first 15 sec of flight.

The Nike Cajun System—A series of seven Nike Cajun firings was conducted at White Sands Missile Range by the U. S. Army Signal Missile Support Agency during the summer of 1958. Instrumentation in the Nike Cajun (Fig. 10) nose cones included a DPN-41 radar beacon, used in certain propagation studies, and a smoke generator, which was being tested as a wind measurement technique. The smoke trail did not prove satisfactory in this firing, owing to payload limitations and the high speeds involved in reaching a peak altitude in excess of 100 miles.

The Nike Cajun, a relatively large rocket, presents several problems in handling and preparation. Suitable equipment is required to place the various components on the launcher, and a great deal of care must be exercised to assure proper mating of the component parts. The Nike Cajun is launched from a standard Nike Ajax launcher. The booster is fitted with a set of four fins which are considerably strengthened to stand the increased acceleration resulting from the small load. The booster burns for 3 sec, attaining a speed of 3200 ft sec. After booster burnout, aerodynamic drag separates the stages, and the Cajun coasts for approximately 15 sec. The Cajun motor then ignites, burning for 3 sec and then coasting in a trajectory that carries it well above 100 miles. The record altitude attained with a 67-lb payload is 121 miles Excellent trajectory data were obtained by radar tracking of the DPN-41 transponder.

Conclusions—The rocket firing crews of the U. S. Army Signal Missile Support Agency have fired a total of 120 meteorological rockets; 10 of them performed satisfactorily, and data wer obtained from 73 of the firings. The rather lo percentage of completely successful firings partly due to testing of experimental components. Instances in which a series of firing

FIG. 10—The Nike Cajun rocket during launchin preparations.

as been conducted with tested hardware has esulted in successful firings in excess of 75 per ent. Further improvement in rocket and sensor eliability can be expected as experience inreases and production problems are minimized.

The training of rocket launching crews, contruction of launchers, perfection of radar trackng techniques, and selection of data handling echniques have been pursued by personnel of e U. S. Army Signal Missile Support Agency ith the intention of obtaining data at high ltitudes on a systematic basis. This aim is now aking form in the planned six-station rocket etwork first proposed several years ago by Dr. ans aufm Kampe of the U. S. Army Signal esearch and Development Laboratories. This etwork is expected to be composed of stations t Wallops Island, Virginia; Patrick Air Force ase, Florida; White Sands Missile Range, ew Mexico; Point Mugu, California; Fort reely, Alaska; and Fort Churchill, Canada.

The initial test will involve daily firings from ch of the six stations for a period of 1 month uring each season of the year. The data obined will provide information about the derability of such measurements and will point t seasonal effects that might prove important meteorologists. The data will be reduced in a rm compatible with standard balloon observaons and will be made available to all interested gencies.

Despite the many disadvantages of the rocket stem, it can be expected to open new areas of terest for the meteorologist by obtaining easurements in this relatively unexplored reon.

References

Anderson, A. D., and W. E. Hoehne, Experiments using window to measure high altitude winds, *Bull. Am. Meteorol. Soc., 37,* 454–457, 1956.

Anderson, A. D., Comments on upper air wind measurements with small rockets, *J. Meteorol., 14,* 473–474, 1957.

aufm Kampe, H. J., Upper air wind measurements with small rockets, *J. Meteorol., 13,* 601–602, 1956.

Battan, L. J., Use of chaff for wind measurements, *Bull. Am. Meteorol. Soc., 39,* 258–260, 1958.

Cline, D. E., Rocket beacon wind sensing system, *U. S. Army Signal Research and Development Laboratory, Fort Monmouth, New Jersey, Eng. Rept. E-1205,* 1957.

Haig, T. O., and Vincent E. Lally, Meteorological sounding systems, *Bull. Am. Meteorol. Soc., 39,* 401–409, 1958.

Jenkins, K. R., and W. L. Webb, Rocket wind measurements, *U. S. Army Signal Missile Support Agency, White Sands Missile Range, New Mexico, Tech. Memo. 531,* 1958.

Mervill, H. J., Structure of the wind, *U. S. Army Signal Research and Development Laboratory, Fort Monmouth, New Jersey, Tech. Memo. M-1195,* 1949.

Spencer, N. W., Forty-seven rocket firings at Fort Churchill, ARS preprint 635–58, 1958.

Stroud, W. G., Meteorological rocket soundings in the arctic, *Jet Propulsion* (American Rocket Society), *28,* 817–822, 1958.

Thaler, W. J., and J. E. Masterson, A rapid high altitude wind determining system, abstract in *Bull. Am. Meteorol. Soc., 37,* 177, 1956.

Vaughn, W. W., Use of chaff to determine wind velocity at aircraft altitude, *U. S. Air Force Armament Center, Eglin Air Force Base, Florida, Tech. Publ.,* 1957.

Webb, W. L., K. R. Jenkins, and G. Q. Clark, Flight testing of the Arcas, *U. S. Army Signal Missile Support Agency, White Sands Missile Range, New Mexico, Tech. Memo. 623,* 1959.

urnal of Geophysical Research Volume 64, No. 11 November, 1959

The Use of a Radar Beacon for Telemetering Precipitation Data

Dewey R. Soltow and Richard D. Tarble

U. S. Weather Bureau
Washington 25, D. C.

Abstract—In many regions of the United States, the lack of observers and other factors prevents the obtaining of precipitation reports from areas repeatedly subjected to flooding. To overcome these difficulties it has been necessary to resort to remote telemetering devices to obtain hydrologic data, such as precipitation, river stage, and tide stage. A method of using the radar beacon for collecting and displaying tipping-bucket rain-gage data on the *PPI* scope of a radar is presented in this paper. Applications of this beacon to other hydrologic data collections are indicated.

Introduction—One of the major problems in veloping an adequate flood warning system, rticularly in remote headwater areas, has been e obtaining of prompt reports. Headwater sins located in mountainous regions suffer om a lack both of observers and of reliable mmunications.

Many groups have been working on the delopment of radio telemetering devices to llect precipitation data from these inaccessible eas. This development, as well as procureent, installation, and maintenance of these dio rain gages, has proved to be extremely stly. As a result, very few have been successlly placed into operation.

The past few years have been marked by a pid increase in the number of radar sets stalled for weather use. By the end of 1958, e U. S. Weather Bureau had in operation 68 nverted airborne AN/APS-2 radars (designted WSR-1 and 3) exclusively for meteorogic and hydrologic purposes. In addition, the st of 31 new and more powerful radar sets VSR-57) has been placed in operation at iami, Florida, with the remainder to be inalled during 1959 and 1960.

This increased use of radar is resulting in ntinuously improving surveillance of the disbution and intensity of rainfall, especially in eas where precipitation gages are widely parated. Though an indication of the varian in intensity of precipitation can be dermined by radar, no completely satisfactory lationship has yet been established between the intensity of the radar echo and recorded precipitation. Therefore, it is necessary to verify the echo intensity through precipitation reports from the area where echoes are detected. The present rain-gage network is too sparse to be satisfactory, and radio-telemetering is too expensive.

One solution to the problem of obtaining precipitation data from remote locations exploits the principle of the radar beacon as a telemetering device for use in conjunction with a weather-search radar system. The input of the telemetering circuit of the beacon can be controlled by many meteorological measuring instruments. In the initial stages of development of the radar beacon gaging device, the tipping-bucket rain gage was used.

Beacon operation—The radar beacon itself is a relatively simple electronic device which has been used extensively for identifying planes in flight and as a homing device.

The beacon receiver is interrogated by the pulse signal from the radar. Upon reception of this pulse, the receiver is blanked or switched to an inoperative status, and the received pulse triggers the beacon transmitter output. The beacon transmitter then sends back a pulse with high peak power to the radar receiver on the same frequency as the original radar interrogating pulse. The received beacon signal is displayed as a reinforced target signal or echo on the radar scope display (*PPI* or *A* scope) at the range and azimuth of the beacon (Fig. 1). This echo is generally strong enough to be

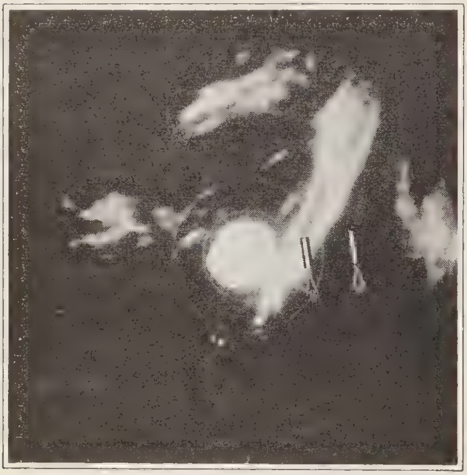


FIG. 1—Report from the radar-beacon rain gage as it appears on the radar scope. The intense echo (outlined in black) shown within the rain gage and nearest the center of the scope is the initial signal from the beacon, indicating its exact location. The second intense echo is the report of the amount of rain, in this instance 0.60 inch. The amount of rain is determined by the distance between the two echoes (10 miles on scope may equal 0.20-inch precipitation).

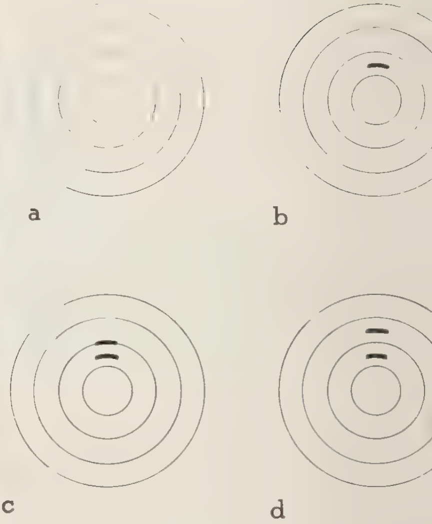


FIG. 2—Appearance of radar scope during op
tion of radar beacon; (a) no precipitation;
initial 'blip' with transmitter activated at pr
termined precipitation accumulation (Σ =
inch); (c) secondary blip appears after assig
increment has accumulated (Σ = 0.40 inch);
blip moves out radially along azimuth line u
successive accumulations (Σ = 0.60 inch).

detected even though heavy precipitation echoes are present in the area of the beacon response. The beacon echo may be seen on the scope face if the receiver gain of the antenna is reduced to the point where the signal strength of the beacon exceeds that of the precipitation echo. This sequence of received radar pulse and transmitted beacon pulse is repeated for every interrogating pulse from the radar.

The initial triggering of the beacon, and thus the initial appearance of the 'blip' on the scope, is dependent upon the accumulation of a predetermined amount of precipitation in a rain gage. In the illustration shown (Fig. 2), this amount is given as 0.20 inch, but it may be any preset value. As each successive predetermined increment of precipitation accumulates, the retransmitted pulse from the beacon is delayed sufficiently to cause the blip to appear at an increased range interval. The movement of the blip radially outward on the radar scope can then be calibrated with precipitation accumulation. When the blip has traversed the interval between its initial location and the edge of
scope, the beacon automatically recycles
the blip returns to its initial location. Additi
increments of rainfall will again cause the
to move outward.

The radar observer can note the time
appearance of the initial blip, as well as s
cessive increments. If these data are plotted
a simple time-versus-amount chart, a ro
approximation of intensity can be obtained

The beacon will continue to transmit rai
amounts until some predetermined time of
(usually set to coincide with standard obse
tion time of cooperative observers in the ar
At this time the keyer is automatically era
and the beacon recycles itself and starts o
If rainfall has ceased, the beacon will shut

During periods when precipitation is
recorded, the beacon automatically turns on
a period of about one hour per day, thus gi
a day-to-day check of the equipment's op
tional status.

The beacon operates on standard 110-

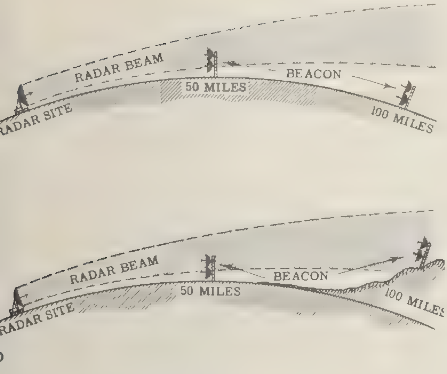


Fig. 3—Effect of terrain on beacon signal re-ption; (a) on level terrain, signal may be re-ived from beacon at 50 nautical miles, none om 100 nautical miles; (b) in mountainous ter-in, signal from beacon may be received from th 50 and 100 nautical-mile range if elevation distant beacon remains in line of sight of radar ation.

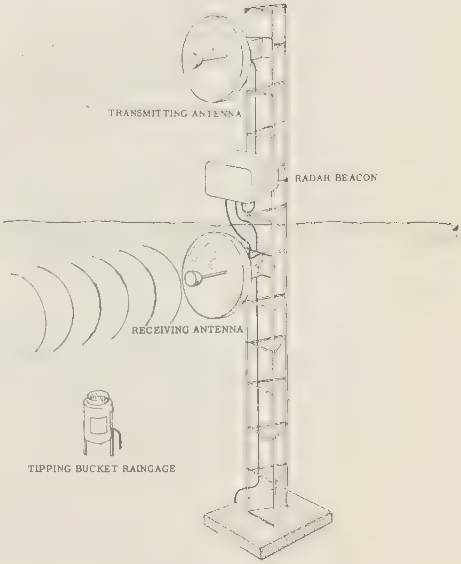


Fig. 4—Beacon system.

lt, 60 cps power. Its relatively low power quirement makes operation feasible with bat-ries charged by wind generators or solar bat-ries.

Operational limitations—At present it is be-eved that the beacon will be limited to line-of-ght operations from the radar station (Fig. 3). ield tests using a slightly larger antenna may veal that the range can be extended slightly eyond the horizon.

At present, the beacon is designed to operate ith a tipping-bucket rain gage. This will limit peration of the beacon to periods when liquid recipitation only is occurring. The changeover a weighing-type gage can easily be effected ith only minor design changes.

Operational advantages—This beacon system as distinct advantages over the conventional dio system. Because it transmits on the same equency as the interrogating radar, there is no ed to obtain a special radio-frequency alloca-on. Radar frequencies also are not affected by terference from the usual atmospheric dis-rbances which so often plague reception of gnals on the lower-frequency bands.

Cost is of great importance when considering e installation of automatic reporting gages. he beacon gage currently costs about $3,000 us installation. The relatively simple electronic circuitry of the beacon lends itself well to miniaturization. The present model occupies a space of approximately 4 cu ft. Through the use of transistors, this space could conceivably be reduced by half, and power requirements would also be reduced. An instrument of this size can easily be installed on existing towers, or, if necessary, it can be installed on a specially constructed tower (Fig. 4) at an additional cost of about $500 per station. Inasmuch as this beacon is limited to line-of-sight, additional cost for repeater stations may be necessary for certain installations. No additional costs for housing the beacon or for a receiver station are necessary.

Maintenance costs should be small. Experience has shown that the beacon will need little attention after installation. The automatic daily operational check will notify the observer of any malfunctioning of the equipment.

Additional applications—The radar beacon can also be used to interrogate recording instruments other than precipitation gages. Such information as snow depth and water equivalent, river stage, tide height, and wind speed and direction, as well as other meteorological observations, could be measured and retransmitted with only minor redesign of the equipment.

Conclusions—Where a radar is available, it is believed that this beacon, though still in the developmental stage, offers substantial advantages over conventional radio-reporting systems in the field of hydrologic and meteorologic data collection. Its low cost, ease of installation, inexpensive maintenance, lack of interference, and —most important—the immediate direct reading of the observation from the face of the radar scope, make this gage system unique in its ability to obtain data. River and flood forecasters, dam operators, and others in the field of hydrology will find a gage of this sort particularly useful in that it facilitates the determination, by radar, of the intensity and the areal distribution of precipitation.

References

Bigler, S. G., and R. D. Tarble, *Applications Radar Weather Observations to Hydrolog* Texas A and M Research Foundation, Colle Station, Texas, 46 pp., 1957.

Hiser, H. W., L. F. Conover, and H. V. Senn, *I vestigation of Rainfall Measurement by Rad* Final Report to U. S. Weather Bureau, Contra Cwb-9012, University of Miami, Coral Gabl Fla., 67 pp., 1957.

Neill, J. C., *Analysis of 1952 Radar and Ra Gage Data,* Report of Investigation No. State Water Survey Division, Urbana, Ill., pp., 1953.

Turner, L. A., Radar Beacons, chp. 8, *Radar Sy tem Engineering,* Vol. 1 Radiation Lab. Seri L. N. Ridenour, ed., McGraw-Hill, New Yo 243–270, 1947.

(Manuscript received February 3, 1959; revis August 20, 1959.)

Conformal Projection of an Ellipsoid of Revolution When the Scale Factor and Its Normal Derivative Are Assigned on a Geodesic Line of the Ellipsoid[1]

MICHELE CAPUTO[2]

Institute of Geophysics
University of California
Los Angeles 24, California

Abstract—In this paper a method of finding the scale factor and the correspondence equations of the conformal projection of an ellipsoid of revolution on a plane is given in the case where the scale factor and its normal derivative are assigned on a geodesic line of the ellipsoid. In particular the method is applied to the case for which the scale factor is constant and has zero normal derivative on that line.

This method is based on the a priori determination of the scale factor, by means of the Souslow-Marussi equation, and by the use of it in the solution of the partial differential equations of the projection.

INTRODUCTION

The recent, widespread developments in nodern practical geodesy, radio navigation, and adio-guided missiles call for new studies in heoretical geodesy, including the field of map rojection. This paper is intended as a contriution to these researches; it gives a method f finding the scale factor (a priori) and the orrespondence equations for the conformal proection $\bar{\Omega}$ of an ellipsoid of revolution on a plane here the scale factor and its normal derivative re assigned on any geodesic line of the ellipsoid. n particular the method is applied to the pecial case in which the scale factor is constant nd has zero normal derivative on that geodesic ne (projection $\bar{\Omega}_0$).

Differential equations for the scale factor and or the correspondence formulae in a generic rojection of one surface on another are first iven in tensor notation, and then the most ecent advances in the modern theory of proection are considered. These advances arise from paper by *Marussi* [1957] in which he derived n intrinsic differential equation relating the inear and two-dimensional scale factors to the urvatures of the two surfaces.

Special attention is given to conformal projections and, in particular, to the conformal projection of an ellipsoid of revolution on a plane. A geodesic coordinate system is introduced on the ellipsoid, and the associated metric tensor is obtained by solving the well-known differential equation relating the Riemann tensor to the curvature of the surface.

In that coordinate system the scale factor is determined a priori by solving *Marussi's* equation [1951a, 1951b, 1957] in the appropriate form for the conformal projections $\bar{\Omega}$. The correspondence formulas are then obtained by considering the equation which relates the known scale factor to the partial derivatives of the correspondence formulas and then associating this equation with the system of differential equations of the projection. This method has already been used by the author [1956–1957] in the case in which the geodesic line, on which values of the scale factor and its normal derivative are assigned, is a meridian of the ellipsoid.

REPRESENTATION OF ONE SURFACE ON ANOTHER

Symbols and conditions—Let Σ and $\bar{\Sigma}$ be the assigned surfaces of curvatures K and $\bar{K}$, and let x^1, x^2, $\bar{x}^1$, $\bar{x}^2$ be the curvilinear coordinate systems on them. Also let a_{ij}, $\bar{a}_{ij}$ be the associated first fundamental metric tensors, and let ds, $d\bar{s}$ be the elementary displacements on the two surfaces. Then

[1] Publication No. 142, Institute of Geophysics, Jniversity of California.

[2] On leave from University of Trieste.

$$ds^2 = a_{ij}\ dx^i\ dx^j \text{ and } d\bar{s}^2 = \bar{a}_{ij}\ d\bar{x}^i\ d\bar{x}^j \qquad (1)$$

are two positive definite quadratic forms.

Now let

$$\bar{x}^i = \bar{x}^i(x^i, x^j) \quad \text{and} \quad i, j = 1, 2 \qquad (2)$$

be the equations of a generic projection Ω of Σ on $\bar{\Sigma}$; differentiating equation (2), substituting in (1), and putting $a^*{}_{ij} = \bar{a}_{hk}\bar{x}_{/j}{}^h\bar{x}_{/i}{}^k$, we have

$$d\bar{s}^2 = \bar{a}_{ij}\bar{x}_{/h}^{\ i}\ \bar{x}_{/k}^{\ j}\ dx^h\ dx^k = a_{hk}{}^*\ dx^h\ dx^k \quad (3)$$

If with the transformation (2) we refer $\bar{\Sigma}$ to the coordinates x^i, then $a^*{}_{ij}$ is the new metric tensor on $\bar{\Sigma}$. In this system the couples of points on Σ and $\bar{\Sigma}$, corresponding in Ω, have the same coordinates x^i.

Later we shall have need for the relation between the Christoffel symbols of the second kind on the two surfaces. The Christoffel symbols of the second kind on $\bar{\Sigma}$ in the metric of tensor $a^*{}_{ij}$ are

$$\overline{\begin{Bmatrix} & k & \\ i & & j \end{Bmatrix}} = \frac{a^{*hk}}{2}\left(\frac{\partial a^*_{hi}}{\partial x^j} + \frac{\partial a^*_{jh}}{\partial x^i} - \frac{\partial a^*_{ij}}{\partial x^h}\right) \quad (4)$$

Let us replace partial derivatives by the covariant ones in the second member of (4).

$$\frac{a^{*hs}}{2}(a_{si/j} + a_{js/i} - a_{ij/s}) \\ = \overline{\begin{Bmatrix} & k & \\ i & & j \end{Bmatrix}} - \begin{Bmatrix} & k & \\ i & & j \end{Bmatrix} \qquad (5)$$

Equation (5) is a tensor which characterizes the difference between the connections of Σ and $\bar{\Sigma}$, and we will indicate it by $T_{ij}{}^k$; its product with $a_{ij}{}^*$ is a vector τ of components $\tau^s = a^{*ij}T_{ij}{}^s$.

Let us consider on Σ a displacement $dP = t\ ds$ of components dx^i, where t is the unit tensor and ds the modulus of dP. The covariant components of t are

$$\lambda^i = \frac{dx^i}{ds}$$

If the two displacements (1) correspond in Ω, the scale factor m_t of Ω in the direction t will then be

$$m_t{}^2 = \frac{d\bar{s}^2}{ds^2} = \frac{a^*_{ij}\ \ dx^i\ dx^j}{ds^2} = a_{ij}{}^*\lambda^i\lambda^j \quad (6)$$

The projection $\Omega'$is defined when we give generic couple of functions (2) with Jacobia not equal to zero. The scale factor, howeve cannot be given arbitrarily a priori; it mus satisfy the equation [*Marussi*, 1957]

$$U\bar{K} = 2K - \Delta_2\,\sigma + \operatorname{div}\tau \\ + \tau \operatorname{grad}\sigma - a^{ir}T_{hr}{}^jT_{ij}{}^h \qquad (7$$

where σ is the logarithm of the two-dimensiona scale factor;

$$U = a^{*ij}a_{ij} = m_1{}^2 + m_2{}^2 \qquad (8$$

is the sum of the two principal scale factors and the curvatures K and $\bar{K}$ are computed i the corresponding points.

We shall use (7) later to find the scale facto of the conformal projection of an ellipsoid o revolution on the plane with the boundar condition requirement that the scale factor i assigned together with its normal derivative o a geodesic of the ellipsoid.

Conformal projections—In the case of a conformal projection, m_t must not depend on th direction of t, and (6), which can also be writte

$$m_t{}^2 = \frac{a_{ij}{}^*\ dx^i\ dx^j}{a_{is}\ dx^i\ dx^j} \qquad (9$$

shows that in this case the matrix

$$\begin{bmatrix} a_{ij}{}^* \\ a_{ij} \end{bmatrix} \qquad (1$$

must have minors with unit characteristic Introducing μ as the natural logarithm of we find that (6) is then

$$a_{ij}m^2 = a_{ij}e^{2\mu} = a_{ij}{}^* \qquad (1$$

In this case it is also well known that

$$T_{ri}{}^i = \sigma_{/r} = 2\mu_{/r} \quad \text{and} \quad \tau^i = 0 \quad (1$$

and with (11) and (12) substituted in ([*Marussi*, 1951a, 1957]

$$\bar{K}e^{2\mu} = K - \Delta_2\mu \qquad (1$$

This is the equation that conditions the sca factor to the curvature on the correspondi points of the two surfaces in the conform projections. It will be used later to obtain t scale factor in the projection $\bar{\Omega}$.

The condition for (10) also gives the syst f partial differential equations for a conformal rojection. In fact, going back to the coordinates ystem x^i and $\bar{x}^i$ on Σ and $\bar{\Sigma}$, we see that that ondition can be formally expressed by saying hat the matrix

$$\begin{bmatrix} \bar{a}_{hk} \dfrac{\partial \bar{x}^h}{\partial x^i} \dfrac{\partial \bar{x}^k}{\partial x^j} \\ \\ a_{ij} \end{bmatrix} \tag{14}$$

as minors with unit characteristics, which gives he system of partial differential equations for conformal projection of Σ on $\bar{\Sigma}$.

If the coordinates systems x^i and $\bar{x}^i$ are rthogonal, then $a_{12} = \bar{a}_{12} = 0$, and (14) can e written

$$\frac{\partial \bar{x}^1}{\partial x^1} = \pm\sqrt{\frac{a_{11}\bar{a}_{22}}{a_{22}\bar{a}_{11}}}\,\frac{\partial \bar{x}^2}{\partial x^2}$$
$$\frac{\partial \bar{x}^2}{\partial x^2} = \mp\sqrt{\frac{a_{11}\bar{a}_{11}}{a_{22}\bar{a}_{22}}}\,\frac{\partial \bar{x}^1}{\partial x^1} \tag{15}$$

t is useful to notice that (11) in this case is

$$n^2 = \frac{\bar{a}_{22}}{a_{22}}\left(\frac{\partial \bar{x}^2}{\partial x^2}\right)^2 + \frac{\bar{a}_{11}}{a_{22}}\left(\frac{\partial \bar{x}^1}{\partial x^2}\right)^2$$
$$= \frac{\bar{a}_{11}}{a_{11}}\left(\frac{\partial \bar{x}^1}{\partial x^1}\right)^2 + \frac{\bar{a}_{22}}{a_{11}}\left(\frac{\partial \bar{x}^2}{\partial x^1}\right)^2 \tag{16}$$

The particular case where $\bar{\Sigma}$ is a plane and $^1 = y$, $\bar{x}^2 = x$ is a Cartesian orthogonal system ill be considered later. System (15) and equation 16) are then

$$\frac{\partial x}{\partial x^2} = \pm\sqrt{\frac{a_{22}}{a_{11}}}\,\frac{\partial y}{\partial x^1}\,,$$
$$\frac{\partial y}{\partial x^2} = \mp\sqrt{\frac{a_{22}}{a_{11}}}\,\frac{\partial x}{\partial x^1} \tag{17}$$

nd

$$\imath^2 = \frac{1}{a_{22}}\left[\left(\frac{\partial y}{\partial x^2}\right)^2 + \left(\frac{\partial x}{\partial x^2}\right)^2\right]$$
$$= \frac{1}{a_{11}}\left[\left(\frac{\partial x}{\partial x^1}\right)^2 + \left(\frac{\partial y}{\partial x^1}\right)^2\right] \tag{17$'$}$$

Geodesic coordinates on the ellipsoid of revoluon—One of our purposes in this paper is to find e scale factor and the equations of the proction $\bar{\Omega}$ in the case where Σ is an ellipsoid of volution, $\bar{\Sigma}$ is a plane, and $\bar{\Omega}$ is a conformal projection of Σ on $\bar{\Sigma}$ which has the scale factor m and its normal derivative assigned on a geodesic line γ^* of Σ.

These conditions with equation (13) determine μ (and m). Since m is known, system (17) will then give us the correspondence formulas of $\bar{\Omega}$.

Let us now introduce on Σ a system of geodesic coordinates. Let P_0 be a point of Σ and let γ^* be a geodesic of asimuth z^* in P_0. Every point P, suitably restricted in a region of the surface Σ, which will be considered throughout the following considerations, gives with P_0 a geodesic line γ. Let z_0 be the azimuth of γ in P_0, and let σ be the distance $P\ P_0$ in γ. Now let $\alpha = z_0 - z^*$; α and σ are geodesic coordinates of P with respect to the pole P_0 and to the fundamental geodesic γ^*.

The following notation will be used.

φ	latitude on Σ
ρ, r	radii of curvature of the meridians and parallels on Σ
z, z^*, z_0	azimuths of geodesic lines
$K = \cos\varphi/r\rho$	curvature of Σ
a, b, e	semimajor, semiminor axis, and eccentricity of Σ

It is well known that the system of geodesic coordinates is orthogonal and the displacements have the form

$$ds^2 = d\sigma^2 + g^2(\alpha, \sigma \mid \varphi_0)\,d\alpha^2 \tag{18}$$

where φ_0 is the latitude of P_0 and $g(\alpha, \sigma \mid \varphi_0)$ is a function which will be determined by solving the well-known equation relating the curvature of Σ to the determinant of the metric,

$$\frac{\partial^2 g}{\partial \sigma^2} = -gK = -g\,\frac{(1 - e^2 \sin^2 \varphi)^2}{a^2(1 - e^2)} \tag{19}$$

For that purpose let us develop K in power series of σ along a geodesic of azimuth z_0 in $P_0(\varphi_0)$

$$K = \sum\nolimits_i \left(\frac{\partial^i K}{\partial \sigma^i}\right)_{P=P_0} \frac{\sigma^i}{i!} \tag{20}$$

where $(\partial^i K/\partial\sigma^i)_{P=P_0}$ will be functions of P_0 and z_0.

Applying Clairaut's theorem on γ, we find that

$$r \sin z = r_0 \sin z_0 = \text{const} = C,$$
$$r_0 = r(\varphi_0) \tag{21}$$

where z is the azimuth along γ. Putting $aV = r_0$, we obtain from (21)

$$\sin^2 \varphi = \frac{\sin^2 z - V^2 \sin^2 z_0}{\sin^2 z - e^2 V^2 \sin^2 z_0}$$

and

$$K = \frac{(1 - e^2) \sin^4 z}{a^2(\sin^2 z - e^2 V^2 \sin^2 z_0)^2} \qquad (22)$$

When K is considered along γ, the only parameter varying in (22) is z, and since the well-known relation

$$\frac{dz}{d\sigma} = \frac{\sin \varphi \sin z}{r} \qquad (23)$$

along γ becomes

$$\frac{dz}{d\sigma} = \sqrt{\frac{a^2 \sin^2 z - C^2}{a^2 \sin^2 z - e^2 C^2}} \frac{\sin^2 z}{C}$$
$$= \sqrt{\frac{\sin^2 z - V^2 \sin^2 z_0}{\sin^2 z - e^2 V^2 \sin^2 z_0}} \frac{\sin^2 z}{a V \sin z_0} \qquad (24)$$

we can compute the derivatives of K with respect to σ by means of $\partial K/\partial\sigma$ and $\partial z/\partial\sigma$,

$$\left(\frac{\partial K}{\partial \sigma}\right)_{P=P_0} = 4A_1 \cos z_0$$

and

$$\left(\frac{d^2 K}{d\sigma^2}\right)_{P=P_0} = -4B_1 \cos^2 z_0$$
$$+ \frac{4e^2(1 - e^2)(1 - V^2)}{a^4(1 - e^2 V^2)^4} \qquad (25)$$
$$A_1 = \frac{e^2 V(1 - e^2)(1 - V^2)^{1/2}}{a^3(1 - e^2 V^2)^{7/2}}$$

and

$$B_1 = \frac{e^2 V^2(1 - e^2)(1 + 6e^2 V^2 - 7e^2)}{a^4(1 - e^2 V^2)^5}$$

Taking account of the relation $z_0 = z^* + \alpha$, we can write the series (20) as

$$K = \frac{1 - e^2}{a^2(1 - e^2 V^2)} - 4A_1 \cos (z^* + \alpha)$$
$$+ \left[\frac{4e^2(1 - e^2)(1 - V^2)}{a^3(1 - e^2 V^2)^4}\right.$$
$$\left. - 4B_1 \cos^2 (z^* + \alpha)\right] \frac{\sigma^2}{2} \qquad (26)$$

where terms after the second have been omitte
The determination of K by (26) will give
reasonable accuracy in the solution of o
problem when Σ is the international ellipso

It is now convenient to solve (19) by mea
of the following series expansion for g.

$$g(\alpha, \sigma \mid \varphi_0) = \sum_i g_i(\alpha \mid \varphi_0)\sigma^i \qquad (2$$

Substituting (27) in (13), we have, consideri
only $i = 0, 1, 2, 3, 4, 5,$

$$2g_2 = -K_0 g_0$$
$$12g_4 = -\left(K_0 g_2 + \frac{dK}{d\sigma} g_1 + \frac{d^2K}{d\sigma^2} \frac{g_0}{2}\right)$$
$$6g_3 = -\left(K_0 g_1 + \frac{dK}{d\sigma} g_0\right) \qquad (2$$
$$20g_5 = -\left(K_0 g_3 + \frac{dK}{d\sigma} g_2\right.$$
$$\left. + \frac{d^2K}{d\sigma^2} g_3 + \frac{d^3K}{d\sigma^3} \frac{g_0}{3!}\right)$$

Since we know that

$$[g(\alpha, \sigma \mid \varphi_0)]_{P=P_0} = 0 \quad \text{and} \quad \left[\frac{\partial g}{\partial \sigma}\right]_{P=P_0} =$$

it follows that $g_0(\alpha \mid \varphi_0) = 0$, $g_1(\alpha \mid \varphi_0) =$
From (28) we have

$$g_2 = 0, \qquad g_3 = -\left(\frac{K_0}{6}\right)_{P=P_0}$$
$$g_4 = \frac{A_1}{3} \cos (z^* + \alpha)$$
$$g_5 = D_1 + \frac{B_1}{10} \cos^2 (z^* + \alpha),$$
$$D_1 = \frac{(1 - e^2)(1 + 12e^2 V^2 - 13e^2)}{120a^4(1 - e^2 V^2)^4} \qquad (2$$

A Priori Determination of the Scale Fact and the Equation of Correspondence

Scale factor—The scale factor of the conforn
projection of Σ on the plane in the case whe
it is given with its normal derivative on t
geodesic γ^* of Σ will now be determined
priori by solving equation (13). Assumi
$x^1 = \sigma$, $x^2 = \alpha$, and computing the Laplaci
$\Delta_2\mu$ on Σ in that coordinate system, we see tl

18) is

$$^2 \frac{\partial g}{\partial \sigma} \frac{\partial \mu}{\partial \sigma} + g^3 \frac{\partial^2 \mu}{\partial \sigma^2} - \frac{\partial g}{\partial \alpha} \frac{\partial \mu}{\partial \alpha} + g \frac{\partial^2 \mu}{\partial \alpha^2} = g^3 K$$

nd taking into account equation (19) we have

$$^2 \frac{\partial g}{\partial \sigma} \frac{\partial \mu}{\partial \sigma} + g^3 \frac{\partial^2 \mu}{\partial \sigma^2} - \frac{\partial g}{\partial \alpha} \frac{\partial \mu}{\partial \alpha} + g^2 \frac{\partial^2 g}{\partial \sigma^2} + g \frac{\partial^2 \mu}{\partial \alpha^2} = 0 \qquad (30)$$

f $f(\sigma)$ and $\psi(\sigma)$ are the value of μ and its normal erivative on $\gamma^*(f(\sigma)$, and $\psi(\sigma)$ are functions which can be expanded in powers of σ) he boundary conditions are

$$(0, \sigma \mid \varphi_0) = f(\sigma) \text{ and } \left(\frac{\partial \mu}{\partial \alpha}\right)_{\alpha=0} = \psi(\sigma) \quad (31)$$

Introducing the following series expansion or μ,

$$\mu(\alpha, \sigma \mid \varphi_0) = \sum_i \mu_i(\alpha \mid \varphi_0)\sigma^i \qquad (32)$$

nd substituting it in (30), we obtain for $\mu_i(\alpha \mid \varphi_0)$ he following relations.

$$\begin{aligned} \mu_0'' = 0 \quad &\text{and } \mu_2'' + 4\mu_2 = -6g_3 \\ _1'' + \mu_1 = 0 \ &\text{and } \mu_3'' + 9\mu_3 = -12g_4 \end{aligned} \qquad (33)$$

f we use conditions (31) and making some imple calculations [it is not restrictive to assume hat $f(0) = 0$ which gives $\mu(0, 0 \mid \varphi_0) = 0$ $\imath(0, 0 \mid \varphi_0) = 1]$, we see that

$$\begin{aligned} _0 &= 0 \text{ and } \mu_1 = f'(0) \cos \alpha + \psi'(0) \sin \alpha \\ _2 &= (\tfrac{1}{2}f''(0) + \tfrac{3}{2}g_3) \cos 2\alpha \\ &+ \psi''(0) \sin 2\alpha - \frac{3}{2} g_3 \\ u_3 &= \left(\frac{1}{6} f'''(0) + \frac{A_1}{2} \cos z^*\right) \cos 3\alpha \\ &+ \left(\frac{1}{18} \psi'''(0) - \frac{A_1}{6} \cos z^*\right) \sin 3\alpha \\ &- \frac{A_1}{2} \cos (z^* + \alpha) \end{aligned} \qquad (34)$$

fter further calculation we find

$$m^2 = e^{2\mu} = \sum_i M_i\sigma^i \qquad (35)$$

where

$$M_i = m_0^2 \sum_i \frac{2^{\beta+\gamma+\cdots+\rho}}{\beta!\gamma! \cdots \rho!} \mu_1{}^\beta \mu_2{}^\gamma \cdots \mu_r{}^\rho \qquad (36)$$

$$\beta + 2\gamma + \cdots + r\rho = i$$

Formula (36) will be used below to find the equations of correspondence.

Equations of correspondence—We shall now find the equations of the correspondence $\overline{\Omega}$ by using (17) and (17′), which are

$$\begin{cases} \dfrac{\partial x}{\partial \alpha} = \pm g \dfrac{\partial y}{\partial \sigma} \\ \dfrac{\partial y}{\partial \alpha} = \mp g \dfrac{\partial x}{\partial \sigma} \end{cases} m^2 = e^{2\mu} = \left(\frac{\partial x}{\partial \sigma}\right)^2 + \left(\frac{\partial y}{\partial \sigma}\right)^2 \quad (37)$$

and by assuming that a σ power series is known for m^2

$$m^2 = \sum_i M_i\sigma^i \text{ and } M_i = M_i(\alpha \mid \varphi_0) \quad (38)$$

This system of three equations and only two unknown functions will have a solution because m is a solution of (13) and has the same boundary conditions as for system (17).

In order to solve system (37) let us consider the following power series for (2).

$$x = \sum_i x_i(\alpha \mid \varphi_0)\sigma^i$$

and

$$y = \sum_i y_i(\alpha \mid \varphi_0)\sigma^i \qquad (39)$$

Substituting (39) and (38) into (37) with evident notations, we have the following equations.

$$\begin{gathered} x_n' = \sum^{i+j=n} (i + 1)g_j y_{i+1} \\ y' = -\sum^{i+j=n} (i + 1)g_j x_{i+1} \\ n = 0, 1, 2, \cdots \end{gathered} \qquad (40)$$

$$\begin{gathered} 2n(x_1x_n + y_1y_n) \\ + \sum^{n-1} (ijx_ix_j + hky_hy_k) = M_{n-1} \\ n = 1, 2 \cdots \end{gathered}$$

where Σ^{n-1} is taken for $i + j = h + k = n + 1$; $(i, j, h, k > 1)$. Considering now $n = 0, n = 1$, we have

$$\begin{cases} x_0 = \bar{C}_1 \\ y_0 = \bar{C}_2 \end{cases} \text{and} \begin{cases} x_1 = C_1 \cos\alpha + C_2 \sin\alpha \\ y_1 = -C_1 \sin\alpha + C_2 \cos\alpha \end{cases}$$

$$C_1^2 + C_2^2 = M_0 \qquad (41)$$

Equations (41) contain three arbitrary constants. These can easily be determined by choosing the x, y Cartesian coordinates with the origin 0 at the corresponding point of P_0 which gives $\bar{C}_1 = \bar{C}_2 = 0$, and by choosing the x axis tangent at 0 to the transformed line of the fundamental geodesic γ^*. This choice gives $(\partial y/\partial\sigma)_{P=P_0} = 0$ and $y_1(0) = 0$ in the series (39). Equations (41) are then

$$\begin{cases} x_0 = 0 \\ y_0 = 0 \end{cases} \qquad (42)$$
$$\begin{cases} x_1 = \sqrt{M_0} \cos\alpha \\ y_1 = -\sqrt{M_0} \sin\alpha \end{cases}$$

It is easily seen that the choice of sign in condition $C_1^2 + C_2^2 = M_0$ is unessential and that M_0 must be different from 0 and not dependent on α.

For $n = 2$ we find

$$\begin{cases} x_2' = 2y_2 \\ y_2' = -2x_2 \end{cases} \qquad (43)$$
$$4\sqrt{M_0}\,(x_2 \cos\alpha - y_2 \sin\alpha) = M_1$$

and, if we substitute the third equation in the other two,

$$\begin{cases} x_2' = 2x_2 \cot\alpha - \dfrac{M_1}{2\sqrt{M_0}\sin\alpha} \\[2ex] y_2' = -2y_2 \tan\alpha - \dfrac{M_1}{2\sqrt{M_0}\cos\alpha} \end{cases} \qquad (44)$$
$$4\sqrt{M_0}\,(x_2 \cos\alpha - y_2 \sin\alpha) = M_1$$

The first two are linear first-order differential equations with solutions

$$x_2 = C_2 \sin^2\alpha - \frac{\sin^2\alpha}{2\sqrt{M_0}} \int \frac{M_1}{\sin^3\alpha}\, d\alpha$$
$$\qquad (45)$$
$$y_2 = C_2' \cos^2\alpha - \frac{\cos^2\alpha}{2\sqrt{M_0}} \int \frac{M_1}{\cos^3\alpha}\, d\alpha$$

The constants C_2 and C_2' can be determined by means of the third equation of the system, which must be satisfied for all values of

In the same way, for any n system (39) giv a system of two differential equations and boundary condition equation for the solution which can be reduced to the solution of tw linear first-order differential equations who arbitrary constants can be determined by mea of the boundary condition equation.

Conformal Projection of an Ellipsoid o Revolution on a Plane Which Has the Sca Factor Constant on a Geodesic of th Ellipsoid and Has a Zero Normal Derivati There

The scale factor—By using the procedure the preceding section we can now find the sca factor and the equations of the conformal pr jection of Σ on a plane in the case where t scale factor is constant on the geodesic and i normal derivative is zero there. (This speci projection will be indicated as $\bar{\Omega}_0$). To foll this procedure we must first find the square the scale factor m^2 as power series of σ.

The conditions that the scale factor m mu be 1 on the geodesic γ^* and have a zero norm derivative there gives $f(\sigma) = \psi(\sigma) = 0$; formulas (34) become

$$\mu_1 = \mu_0 = 0 \quad \text{and} \quad \mu_2 = -3g_3 \sin^2\alpha \qquad (4$$

$$\mu_3 = A_1 \sin^2\alpha \left(\frac{\sin z^* \sin\alpha}{3} - \cos z^* \cos\alpha\right.$$

Equations of the projection—To obtain t equations of the correspondence $\bar{\Omega}_0$ we mu consider system (40) where the M_i's are giv in this case by means of (46). Since $M_0 =$ $M_1 = 0$, systems (42) and (45) easily give

$$x_0 = y_0 = x_2 = y_2 = 0$$
$$x_1 = \cos\alpha \qquad (4$$
$$y_1 = -\sin\alpha$$

For $n = 3$, equations (40) are

$$\begin{cases} x_3' = -g_3 y_1 \sin\alpha + 3y_3 \\ y_3' = -g_3 x_1 \cos\alpha - 3x_3 \end{cases} \qquad (4$$

$$6x_3 \cos\alpha - 6y_3 \sin\alpha$$
$$= M_2 = 2\mu_2 = -\sigma g_3 \sin^2\alpha$$

Substituting the third equation in the first a cond in (48) we obtain the following linear rst-order ordinary differential equations with ıe same condition.

$$,' - 3x_3 \cot \alpha = 2g_3 \sin \alpha \qquad (49)$$

$$,' + 3y_3 \tan \alpha = -g_3 \cos \alpha + 3g_3 \frac{\sin^2 \alpha}{\cos \alpha}$$

heir solutions are

$$\begin{aligned} x_3 &= C_3 \sin^3 \alpha - 2g_3 \sin^2 \alpha \cos \alpha \\ y_3 &= C_3' \cos^3 \alpha - g_3 \sin \alpha \cos 2\alpha \end{aligned} \qquad (50)$$

oting that $(\partial\mu/\partial\alpha)_{\alpha=0} = 0$, we find that the ›odesic γ^* must be transformed into a straight ıe (by the theorem of Schols) which must ›incide with the x axis. We must also have $(0) = 0$. Then $C_3' = 0$.

It is easily seen that to satisfy the third equaon in (48) we must have $C_3 = 0$.

In the same way we obtain x_4 and y_4 by ıtting $n = 4$ in (40).

$$\begin{aligned} _4 &= -\frac{A_1}{6} [\sin^2 \alpha(5 - 6 \sin^2 \alpha) \cos z^* \\ &\qquad - 2 \sin^3 \alpha \cos \alpha \sin z^*] \\ _4 &= -\frac{A_1}{6} [\sin 2\alpha(1 - 3 \sin^2 \alpha) \cos z^* \\ &\qquad - \sin^2 \alpha \cos 2\alpha \cos z^*] \end{aligned} \qquad (51)$$

Numerical values—We notice that x_3 and y_3 ›pend only on φ_0 and α, which means that if ıe series are terminated after these terms Σ ›ppears to be a sphere of radius $a\sqrt{1 - e^2/}$ $- e^2 \sin^2 \varphi_0)$. The terms x_4 and y_4 depend ı z_0^* in addition to φ_0 and α. Therefore, if the ries are terminated after the fourth terms, $\bar{\Omega}_0$ cludes the ellipsoidal correction. Upper limits r the numerical values of $|x_3\sigma^3|$, $|y_3\sigma^3|$ for $\sigma = 1610$ km and $\sigma = 3220$ km are given in Table 1. In both cases $|\alpha| < \pi/8$. Upper limits are also given for $|x_4\sigma^4|$, $|y_4\sigma^4|$ with $z^* = \pi/4$. The same values are used for σ and α as before. These upper limits of the terms of (38) give an idea of the accuracy which can be obtained by terminating the series.

TABLE 1—*Upper limits of the terms of the series of the projection* $\bar{\Omega}_0$

σ, km	$\lvert x_3\sigma^3 \rvert$, m	$\lvert x_4\sigma^4 \rvert$, m	$\lvert y_3\sigma^3 \rvert$, m	$\lvert y_4\sigma^4 \rvert$, m
1,610	4,800	7	4,700	6
3,200	38,000	120	33,000	100

Acknowledgment—This research was carried out under Office of Naval Research Contract Nonr 233(19).

REFERENCES

ALLARD, P., La projection conforme de deformation stationnaire le long d'une courbe, *Annales Hydrographiques*, Imprimerie Nationale, Paris, 1948.

CAPUTO, M., Su alcune problemi al contorno della teoria delle rappresentazioni conformi, *Atti ist, veneto sci. lettere ed arti*, 9–21, 1956–1957.

MARUSSI, A., Su alcune proprieta integrali delle rappresentazioni conformi di superfici su superfici, *Rend. accad. nazl. Lincei, 10*, ser. 8, fasc 4, 307–310, 1951a.

MARUSSI, A., Determinazione apriori del modulo di deformazione lineare nella rappresentazione conforme di Gauss, *Rend. accad. nazl. Lincei, 11*, ser. 8, fasc 3–4, 198-201, 1951b.

MARUSSI, A., Sulle rappresentazioni fra superfici definite mediante la forma quadratica che ne determina il modulo di deformazione, *Festschr. C. F. Baeschlin*, 201–210, 1957.

(Manuscript received May 16, 1959.)

Tests of the LaCoste-Romberg Surface-Ship Gravity Meter I[1]

J. C. HARRISON

Institute of Geophysics, University of California
Los Angeles, California

Abstract—Gravity measurements made with a LaCoste-Romberg surface-ship gravity meter on board the M. V. *Horizon* along a 300-mile track off southern California are compared with submarine measurements near the track. The agreement (± 5 mgal) is considered to be as close as can be expected from this type of comparison.

Introduction—The development of the Laoste-Romberg surface-ship gravity meter, toether with some tests made on board the Texas . and M. College ship, the *Hidalgo,* during the ring of 1958, has been described recently by *aCoste* [1959]. The results of these tests were ifficiently good to warrant further, more critil, tests. Accordingly a three-day cruise was ranged for October 29–31, 1958, during which avity measurements were made on board the I. V. *Horizon* along a 300-mile track passing ose to thirteen submarine gravity stations off e coast of southern California. The meter sed was number 5 in the series of LaCosteomberg ship-borne gravity meters, the first ur meters having been built for use on subarines. Meter 5 is the same meter used on rlier tests on board the *Hidalgo,* and was ndly loaned for the *Horizon* tests by L. J. B. aCoste. The *Horizon* is an oceanographic rearch vessel, of 505 gross tons and about 970 ns total laden displacement, operated by the ripps Institution of Oceanography of the Unirsity of California. It has a length of 143 feet, beam of 29 feet, and draws about 14 feet of ater. J. G. Cobb (LaCoste and Romberg Comny), C. O. Alexis (Office of Naval Research) d the author took part in these tests.

Description of the tests—The track of the *orizon* during the tests is shown in Figure 1 gether with the positions of the submarine avity stations used for comparison. All but e of the submarine gravity stations had been tablished by the writer with a Vening Meinesz ndulum apparatus at an earlier date [*Harrison and others,* 1957]. The remaining station, number 188, was taken with a LaCoste-Romberg submarine gravity meter and has not been previously reported. Sea conditions varied from calm to slight-to-moderate swells during the tests.

The surface gravity measurements began at point *A* and were made at a speed of 3.5 knots through the positions of stations 1 and 2. Then the speed of the ship was increased to 8.5 knots and this speed was maintained until the ship was anchored off Catalina Island for the first night. Measurements were resumed next morning at point *C* and a continuous profile was run at 8.5 knots across the continental borderland and the continental slope out to point *G*. Here the course was altered to pass through stations 5 and 123 and the speed of the ship was increased to the full cruising speed of 11.5 knots. A further alteration of course was made at point *H* to begin the return journey. Measurements were discontinued at point *I* because of a minor electrical failure in the gravity meter's electronic reading circuits.

Navigation was excellent to the east of San Clemente Island, and frequent accurate fixes were obtained. No fixes were possible after losing radar contact with San Clemente at point *F* until San Clemente was picked up again on the way back to San Diego. Fortunately the continental borderland is characterized by irregular topography which has been charted in detail, and it was possible to establish the ship's track between point *F* and the continental slope just to the west of station 4 by fitting the ship's fathometer profile to the bathymetric charts. West of the continental slope the bottom top-

[1] Institute of Geophysics Publication No. 147.

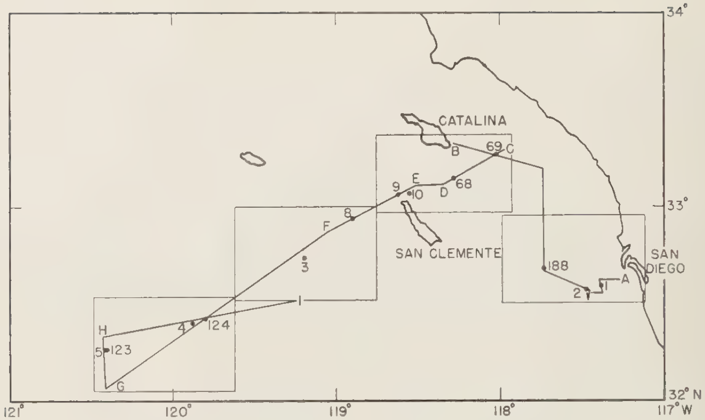


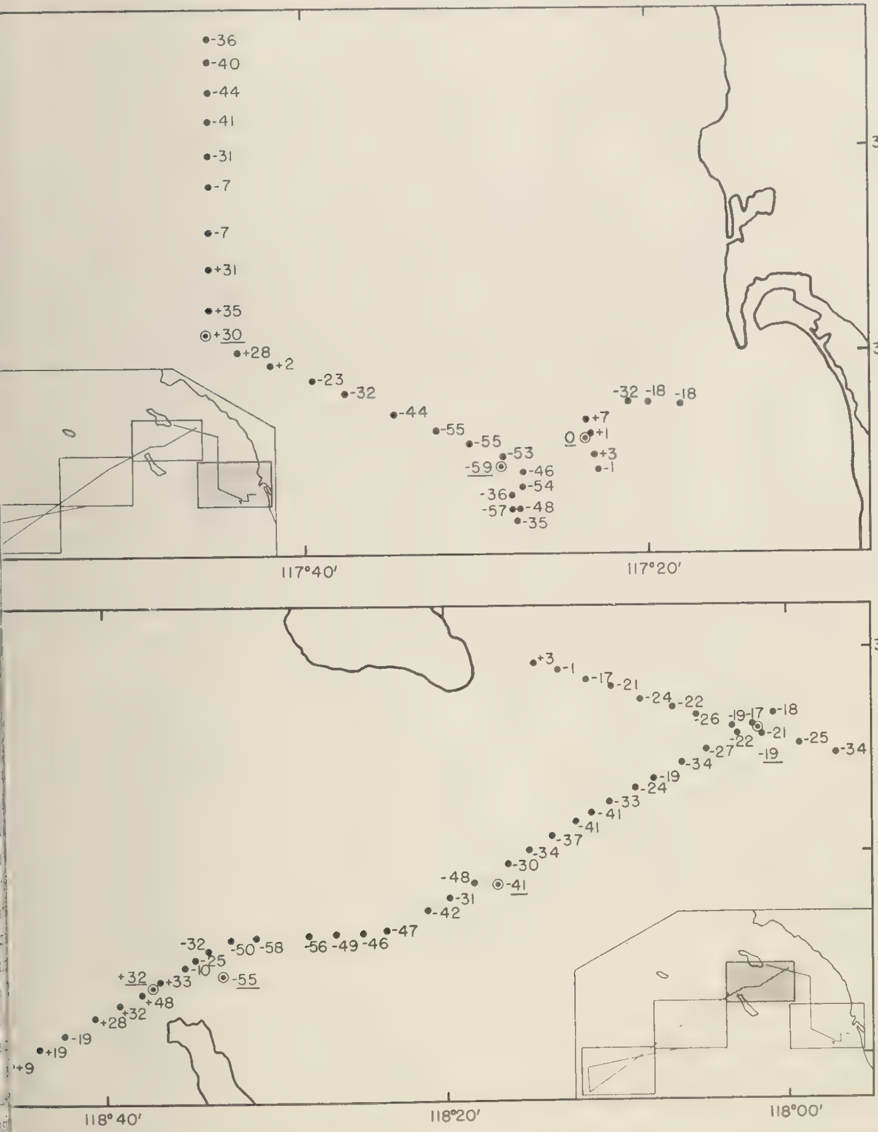
FIG. 1—*Horizon's* track during the tests.

ography is insufficiently detailed to provide a navigational reference system. Although the ship's deep-water fathometer broke down shortly after the change of course at point *H*, echoes were obtained from the shallow-water fathometer when crossing ridges in the continental borderland, and these soundings were sufficient to determine the track between station 4 and point *I*. The estimated accuracy of the ship's speed and position on the portions of the track east of station 4 is 0.25 knots and 0.5 miles, respectively. These uncertainties are doubled west of this station. In these latitudes an error of 1 knot in the east-west component of ship's speed produces an error of about 6.4 mgal in measured gravity, and in a north-south direction an error of 1 mile in position results in an error of 1.4 mgal in the gravity anomaly; therefore the root-mean-square (rms) error introduced into the gravity anomalies by these uncertainties is about 1.6 mgal to the east of station 4 and about 3.2 mgal west of this station.

The gravity measurements—The gravity meter is suspended in gimbals, and it swings to follow the total apparent acceleration. The horizontal accelerations of the instrument are measured, and the second-order correction, which must be subtracted from the total apparent acceleration to obtain the acceleration due to gravity, is automatically computed and applied. In the instrument used in these tests, a small correction proportional to the mean-square vertical acceleration was also necessary. This correction was discovered and its magnitude d
termined during laboratory tests. The mea
square vertical accelerations during the observ
tions at sea were determined from the record
the beam position of the gravity meter. T
need for this second-order correction for vertic
accelerations is not intrinsic in the design of t
meter and it has since been eliminated.

The free-air anomalies inside the areas ou
lined in Figure 1 are shown in Figures 2 and
and the profile *CDEFG* is shown in Figure
Each observation is an average over an interv
of about 10 minutes, though the exact interv
used varied in an unsystematic way between
and 20 minutes. During the runs through st
tions 1 and 2, the corrections for the horizon
accelerations varied between 50 and 220 mg
averaging about 150 mgal. The correspondi
range of rms horizontal acceleration is 10 to
gal, and the average is 18 gal. The rms verti
acceleration averaged about 16 gal (10 m
correction). The horizontal accelerations
creased and the vertical accelerations increas
on altering course to northwest and increas
speed to 8.5 knots after leaving station 2; th
respective rms values became 14 and 20
These disturbing accelerations gradually
creased during the day, until, in the lee
Catalina Island, the rms values of both accele
tions had fallen to around 6 gal. The accelerat
periods on this first day were between 4 an
sec.

Conditions were as calm on the morning



i. 2—Free-air anomalies from surface-ship measurements compared with submarine measurements. Circled points and underlined numbers refer to the submarine measurements.

second day at the start of profile *CDEFG* they had been at the end of the first day, but y gradually became more disturbed as the lter afforded by San Clemente Island was t. The accelerations were at their greatest between station 4 and point *G* and their period increased to 6 to 8 sec. The correction for horizontal acceleration averaged about 200 mgal (20 gal rms horizontal acceleration) and on occasion exceeded 300 mgal. The rms vertical accelera-

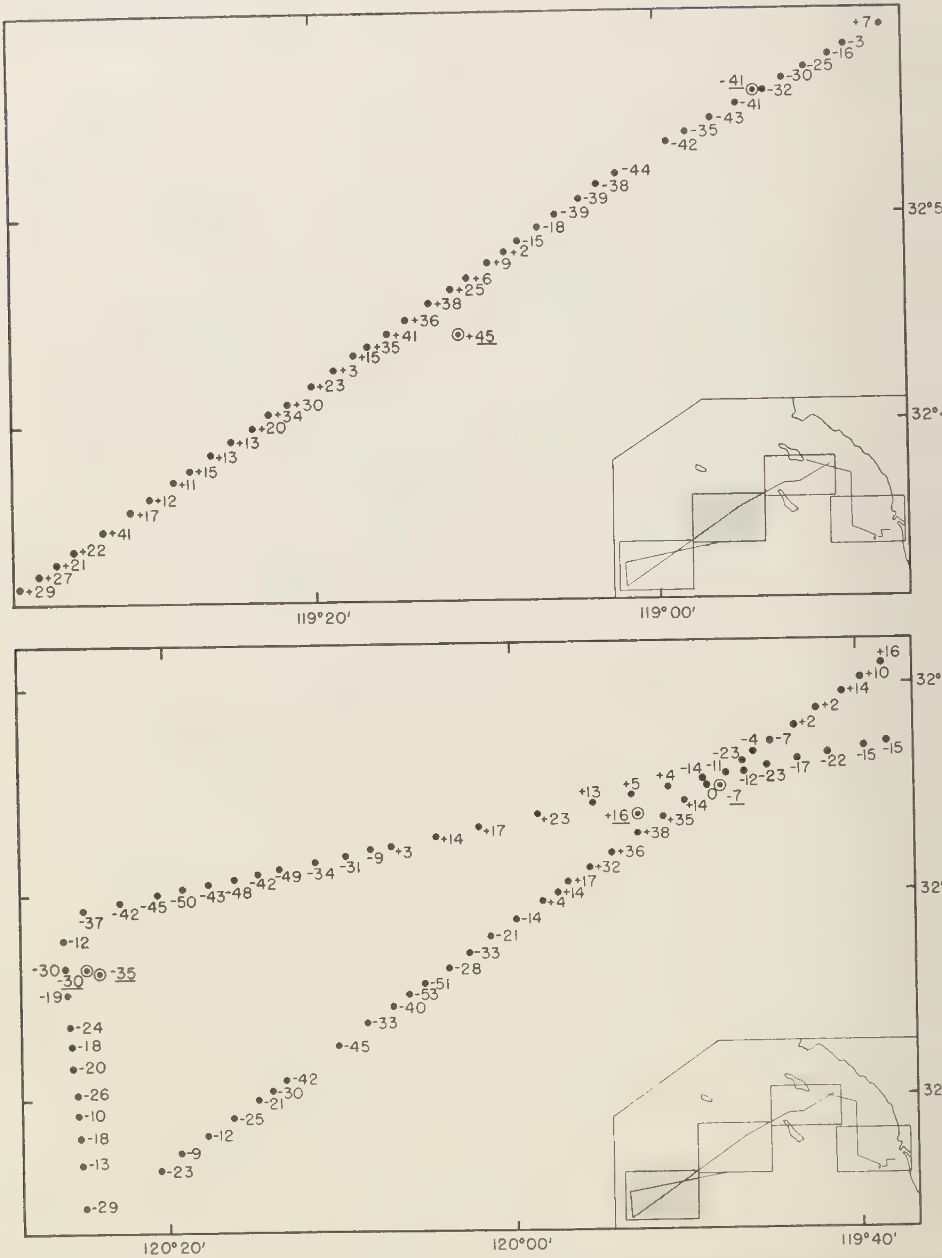


Fig. 3—Free-air anomalies from surface-ship measurements compared with submarine measurem
Circled points and underlined numbers refer to the submarine measurements.

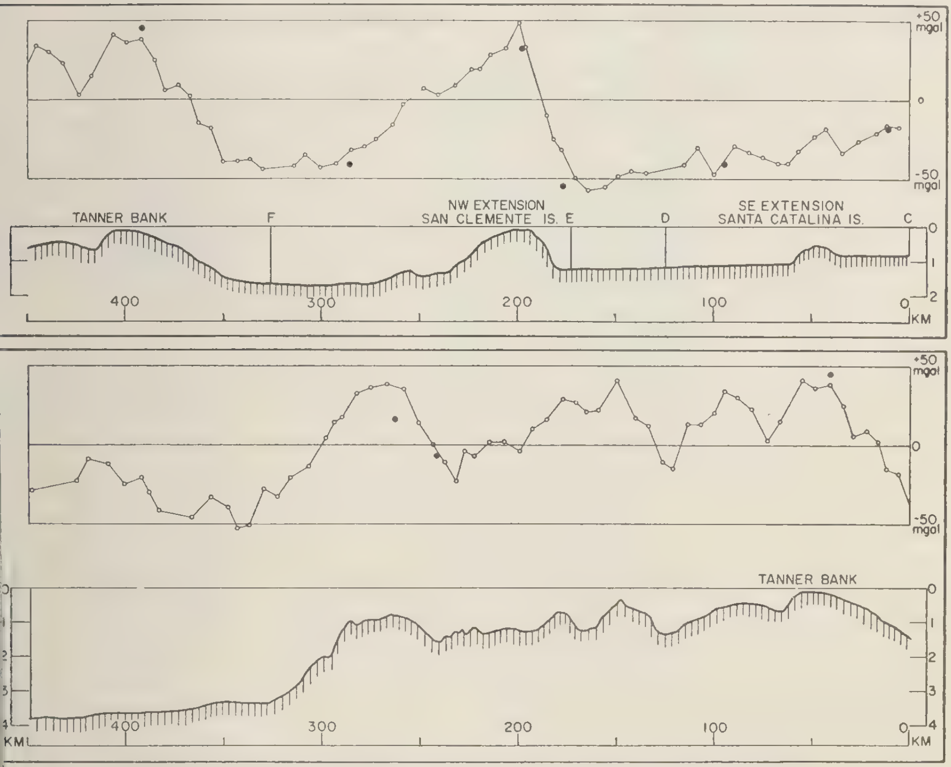


Fig. 4—Topographic and free-air gravity profiles. Solid circles are submarine measurements.

on was also about 20 gal, but the increased eriod of these accelerations rendered the dampg of the gravity meter less effective and the eam hit its stops on five occasions on this poron of the track, the only times it did so on the hole trip. About 3 minutes of observation were st every time the beam hit its stops. The ertical accelerations decreased to an rms value ' about 18 gal and their period dropped to round 5 sec after the change to a northerly urse at point *G*. Conditions quieted still furer after the course change at point *H*; the ns vertical acceleration fell to around 10 gal d the rms horizontal acceleration to about gal. Both accelerations continued to decrease til a leaking capacitor in the gravity meter's otoelectric reading circuit put an end to the servations at point *I*.

It is seen from Figures 2 and 3 that the free-air anomalies often vary rapidly from one measurement to the next. From Figure 4 it is seen that these variations are very closely related to the changes in bathymetry. The close correlation between water depth and free-air gravity anomaly shown in Figure 4 is by itself good reason for believing that the gravity meter readings are valid. Further, the meter was self-consistent. On the two occasions when the *Horizon* crossed its own track there was a difference of only 3 mgal at the crossing near station 69 and 2 mgal near station 124.

It is difficult to make an exact comparison between the submarine and surface-vessel results. The free-air anomaly varies so rapidly that an error of one mile in the position of a submarine station relative to *Horizon's* track can cause an error of more than 10 mgal in the comparison of the gravity anomalies. Both types of meas-

TABLE 1—*Comparisons of surface-ship with submarine gravity measurements*

Station number	Free-air anomaly by submarine measurement, mgal	Free-air anomaly by surface measurement, mgal	Difference, submarine value–surface value, mgal	Notes
1	0	+ 1	−1	
2	−59	−53	−6	
188	+30	+31	−1	
69	−19	−21	+2	Taken on NW course
69	−19	−17	−2	Taken on SW course
68	−41	−42	+1	
10	−55	−32	(−23)	
9	+32	+38	−6	
8	−41	−34	−7	
3	+45	+38	+7	
124	− 7	− 5	−2	Taken on SW course
124	− 7	−13	+6	Taken on E course
4	+16	+21	−5	
5	−30	−30	0	
			−1.1	Mean difference
			±4.4	rms difference.

urement give average values over the duration of each observation, but the submarine measurements were normally taken on a course parallel to the trend of the topography and the surface-ship measurements on a course perpendicular to this. Hence there is a tendency for this averaging process to reduce the amplitude of the peaks and troughs of the gravity profile in the case of the surface-ship measurements. Some interpolation between neighboring surface-ship measurements was necessary in obtaining the values listed in Table 1 and there is some personal judgment involved.

The discrepancy at station 10 is very much greater than that at any other station. The surface-ship values in this vicinity are consistent within themselves and correlate well with the changes in the depth of water as seen from Figure 4, whereas the submarine value, when corrected for the topographical water column, yields a Bouguer anomaly some 25 mgal lower than any in the neighborhood. All these discrepancies would be removed if the true position of station 10 were 1 or 2 miles east of its position in Figure 2. It is always possible that a mistake was made in determining or recording the position of this particular station. The comparison at station 10 is not included in the determination of the mean or rms difference between the two sets of measurements.

The mean difference between the submari and surface-ship measurements is −1.1 mg and the rms difference is 4.4 mgal. This agre ment is as close as can be expected in view the probable error assigned to the submari measurements used in the comparison (+4 mgal for stations 123 and 124, ±2.5 mgal f the remainder) and the uncertainties in maki this type of comparison.

Conclusion—It is possible to obtain measu ments accurate to at least ±5 mgal with t LaCoste-Romberg surface-ship gravity meter calm and slight-swell conditions on board surface vessel of less than 1000 tons total lac displacement without the aid of a stabili platform. Such sea conditions prevail in ma parts of the oceans of the world for a la percentage of the time during certain seas so that this instrument should make it poss for gravity data to be collected at sea m more rapidly and with much greater detail t was possible with submarines. The author had no experience of operating the instrum on larger ships or on ships fitted with stabili fins, but it is reasonable to expect that s ships would provide a satisfactory platform more disturbed sea conditions than could tolerated on the *Horizon*.

As a result of the tests described above, LaCoste and Romberg submarine meter 3 onverted for use on surface-ships. This meter was used on the *Horizon* from February 25 to April 11, 1959, during the first half of an expedition to the Gulf of California. Some 2500 measurements were obtained and more critical ests of the surface-ship meter were made by measuring simultaneously with the surface meter and with an underwater meter which was lowred onto the sea bottom. These tests will be described in a later paper.

Acknowledgments—This work was carried out under contract Nonr 233(19) with the Office of Naval Research. The author is indebted to L. J. B. LaCoste for the loan of the surface-ship gravity meter and to J. G. Cobb and C. O. Alexis for their assistance in taking the measurements.

References

Harrison, J. C., G. L. Brown, and F. N. Spiess, Gravity measurements in the northeastern Pacific Ocean, *Trans. Am. Geophys. Union, 38,* 835–840, 1957.

LaCoste, L. J. B., Surface ship gravity measurements on the Texas A and M College ship, the "Hidalgo", *Geophysics, 24,* 309–322, 1959.

(Manuscript received June 22, 1959.)

A Class of Three-Dimensional Shallow-Water Waves[1]

J. E. CHAPPELEAR

Shell Development Company, Exploration and Production Research Division, Houston, Texas

Abstract—We consider the problem of the calculation of the properties of three-dimensional waves (whose surface profiles have a two-dimensional structure) using the approximations of the shallow-water theory. We find that, if we assume that the first approximation is a uniform flow, there is in the theory a critical speed nearly equal to the square root of the product of the acceleration of gravity and the depth. No steady waves can propagate slower than this velocity. Waves which have this critical velocity are essentially two-dimensional, since they differ from two-dimensional waves only by a steady current. The waves whose velosities are greater than critical may have a wide variety of behaviors, since in this case the velocity potential satisfies a differential equation of the hyperbolic type (in two-space coordinates, not the time) to the second order of approximation. Although we have not been able to construct any solution to the first approximation, other than the uniform flow, we have not been able to find a proof that the uniform flow is a unique solution.

Introduction—The general problem of calculaon of the properties of waves of finite height treated by *Stoker* [1957, ch. 2 and 10] and the iscussion will not be reproduced here. We shall t down the differential equation and the arious boundary conditions, not attempting a etailed motivation or derivation from more eneral considerations. We should like to calulate properties of waves having a three-mensional structure, using the shallow-water pproximation to the equations of hydrodyamics. These waves are to be steady when ewed from a coordinate system translated th a suitable velocity. We should like to escribe the period, depth, and wave height.

Two-dimensional shallow-water waves were vestigated by *Keller* [1948]. He found that litary and periodic waves of finite height were rmally possible. The existence of the solitary ve was shown by *Friedrichs and Hyers* [1954], d the existence of periodic waves by *Littman* 957]. They found that specifying the depth d the height of a solitary wave determined e wave properties completely, whereas the ermination of the periodic wave required further specification of the period.

We present a formal extension of these results three dimensions. In two dimensions, the tical velocity $\sqrt{gd}$, where g is the acceleration of gravity and d the depth of the water, plays an important role. The velocities of the various waves are near to but larger than this critical velocity. We find that this critical velocity is important in the case of three-dimensional waves as well. If we assume that the first approximation is a uniform flow, we find that no waves will propagate below this critical velocity. For a velocity equal to the velocity of the corresponding two-dimensional wave, the only possible generalization of the solution is the addition of a uniform current in the direction parallel to the wave crests. If the velocity is greater than the critical velocity, we also find that the velocity potential satisfies the wave equation; that is, the solutions for ϕ have the property that

$$\phi_y(x, y) = \pm\kappa\phi_x(x, y)$$

or

$$\phi(x, y) = \phi(x \pm \kappa y)$$

It is then possible to have a great number of different solutions. Two particular examples are considered in some detail, one in which the water surface has sinusoidal variations in two directions, and the other in which the surface is flat, except for a single ridge which extends to infinity at both ends. The question about the existence of other waves whose first approximation is not a uniform flow has not been settled here.

Shell Publication No. 198.

List of Symbols

$a(x, y)$	equation (22)
$A(x, y)$	equation (7)
$b(x)$	equation (22)
$B(x, y)$	equation (8)
C	celerity
d	depth
D	equation (17)
$E(x, k)$	elliptic integral of the second kind, equation (39)
$f(x, y)$	equation (11)
$f^+(x, y)$	equation (41)
$f^-(x, y)$	equation (41)
F	$(r_1 - r_3)^{1/2}/2L$
g	acceleration of gravity
$h(x, y)$	equation (13)
H	height of wave
$i(x, y)$	equation (14)
k^2	$(r_2 - r_3)/(r_1 - r_3)$, modulus of $E(x, k)$
l	arbitrary constant, equation (29)
L	$1/Z_0D^{1/2}$
m	arbitrary constant, equation (30)
r_n	roots of polynomial, equation (32)
T	period of wave
x	horizontal coordinate
$\bar{x}$	x/λ
y	horizontal coordinate
$\bar{y}$	y/λ
z	vertical coordinate
$\bar{z}$	z/λ
$Z(x, y)$	surface elevation, $\sum_{n=0}^{\infty} \sigma^n Z_n$
$\bar{Z}$	Z/d
$Z_n(x, y)$	coefficient of σ^n in Taylor series expansion in σ of $Z(x, y)$
η	$\sqrt{D^2/Z_0 - 1}$
κ	arbitrary constant, equation (48)
λ	typical horizontal dimension
λ_0	wavelength
$\phi(x, y, z)$	velocity potential, $\sum_{n=0}^{\infty} \sigma^n \phi_n$
$\phi_n(x, y, z)$	coefficient of σ^n in Taylor series expansion in σ of $\phi(x, y, z)$
$\bar{\phi}(x, y, z)$	$\phi(x, y)/\lambda\sqrt{gd}$
σ	d^2/λ^2

Theory—We employ a coordinate system in which the plane $z = 0$ is the ocean bottom, with the z axis directed vertically upwards. The x and y axes are horizontal. We call the mean water level the plane $z = d$. The free surface is at $z = Z(x, y)$, and the velocity potential is $\phi(x, y, z)$. The coordinate system moves with the velocity C in the x direction, and the motion is steady in this moving system.

The shallow-water approximation depends on the assumption that lengths in the z direction are small in comparison with lengths in the x and y directions. We change the scale of the variables to insure this result. Let λ be a typical length in the z or y direction, and let the depth d be a typical length in the z direction. The wave velocity is assumed to be of order $\sqrt{gd}$. Then if we introduce dimensionless coordinates, the equations to be satisfied are [*Stoker*, 1957, ch. 2] the conservation of mass,[2]

$$\sigma\bar{\phi}_{\bar{x}\bar{x}} + \sigma\bar{\phi}_{\bar{y}\bar{y}} + \bar{\phi}_{\bar{z}\bar{z}} = 0 \tag{1}$$

the boundary condition at the bottom, no flow through the bottom,

$$\bar{\phi}_{\bar{z}} \big|_{\bar{z}=0} = 0 \tag{2}$$

and the coupled boundary conditions on the free surface, Bernoulli's theorem,

$$\begin{aligned}[\tfrac{1}{2}(\sigma\bar{\phi}_{\bar{x}}^{\ 2} + \sigma\bar{\phi}_{\bar{y}}^{\ 2} + \bar{\phi}_{\bar{z}}^{\ 2})& \\ + \sigma\bar{Z}(x, y)]_{\bar{z}=\bar{Z}(x,y)} &= \text{Const.}\end{aligned} \tag{3}$$

and the condition that there is no flow throug the free surface,

$$[\sigma\bar{\phi}_{\bar{x}}\bar{Z}_{\bar{x}} + \sigma\bar{\phi}_{\bar{y}}\bar{Z}_{\bar{y}} - \bar{\phi}_{\bar{z}}]_{\bar{z}=\bar{Z}} = 0 \tag{4}$$

Finally, the gradient of $\bar{\phi}$, the velocity, mus be everywhere bounded. We intend now t expand our solution in powers of σ, which assumed to be small, substitute these expansion into the equations and boundary condition and equate coefficients of the various powe of σ. The equations obtained can then be solv in turn for the unknown functions. We mu assume that, after dropping the bars, $\phi(x, y,$ and $Z(x, y)$ may be expanded in a power ser in σ, since there is no mathematical justificati a priori. There are, however, proofs that t procedure leads to convergent series for vario two-dimensional problems [*Friedrichs and Hye* 1954; *Littman*, 1957]. Let us put

$$\phi(x, y, z) = \sum_{n=0}^{\infty} \sigma^n \phi_n(x, y, z)$$

[2] Note that subscripts are usually employed symbolize partial differentiation.

and

$$Z(x, y) = \sum_{n=0}^{\infty} \sigma^n Z_n(x, y) \qquad (6)$$

We may reduce the number of unknown functions by the use of the differential equation (1), and the boundary-condition equation (2). If we substitute ϕ from equation (5) in equation (1), the term not multiplied by σ is

$$\phi_{0zz} = 0 \qquad (7)$$

The general solution of (7) is obtained by integration. It is

$$\phi_0 = A(x, y)z + B(x, y) \qquad (8)$$

To satisfy (2) we must have

$$A(x, y) = 0 \qquad (9)$$

The next approximation to (1) yields (proportional to σ)

$$\nabla_2{}^2 B + \phi_{1zz} = 0 \qquad (10)$$

where we have used the abbreviation

$$\nabla_2{}^2 B = B_{xx} + B_{yy} \qquad (11)$$

Thus we find as a solution to (10), taking care to satisfy the boundary-condition equation (2).

$$\phi_1 = -(z^2/2)\, \nabla_2{}^2 B(x, y) + f(x, y). \qquad (12)$$

This process can be continued, and we obtain

$$\phi_2 = (z^4/4!)\nabla_2{}^4 B(x, y) - (z^2/2!)\nabla_2{}^2 f(x, y) + h(x, y) \qquad (13)$$

and in the same way

$$\phi_3 = -(z^6/6!)\nabla_2{}^6 B(x, y) + (z^4/4!)\nabla_2{}^4 f(x, y) - (z^2/2!)\nabla_2{}^2 h(x, y) + i(x, y), \text{ etc.} \qquad (14)$$

We now assume that B is a linear function of : and y. This assumption can be verified by ubstituting the lowest order terms of equations 5) and (6) in equations (3) and (4) and retaining vhat is proportional to σ. Physically we have ssumed that the first approximation is a uniform low. We find

$$\tfrac{1}{2}(\nabla_2 B)^2 + Z_0 = \text{Const.} \qquad (15a)$$

and

$$\nabla_2 \cdot (Z_0 \nabla_2 B) = 0 \qquad (15b)$$

It is easily seen that these two equations are satisfied, provided that Z_0 is a constant. The differential equation which B satisfies is obtained by eliminating Z_0 between equation (15a) and equation (15b). The result is

$$\nabla_2{}^2 B[\text{Const.} - \tfrac{1}{2}(\nabla_2 B)^2] + (\nabla_2 B)\cdot(\nabla_2 \nabla_2 B)\cdot \nabla_2 B = 0 \qquad (15c)$$

It is still an open question whether there are additional solutions to (15c), other than the uniform flow.

Without loss of generality, we may choose B to be independent of y. This choice merely aligns the x axis with the direction of propagation of the incoming waves.

We can write

$$\phi(x, y, z) = Dx + \sigma f(x, y) + \sigma^2[-(z^2/2!)\nabla_2{}^2 f + h] + \sigma^3[(z^4/4!)\nabla_2{}^4 f - (z^2/2!)\nabla_2{}^2 h + i] + \cdots \qquad (16)$$

where we have put

$$B = Dx \qquad (17)$$

If we substitute equations (16) and (6) into equations (3) and (4), the terms proportional to σ^2 are

$$Df_x + Z_1 = \text{Const.} \qquad (18)$$

and

$$DZ_{1x} + Z_0 \nabla_2{}^2 f = 0 \qquad (19)$$

Eliminating Z_1, we obtain

$$\left(1 - \frac{D^2}{Z_0}\right) f_{xx} + f_{yy} = 0 \qquad (20)$$

The appropriate boundary conditions are that f_x and f_y are everywhere bounded. The character of the solution is determined by whether D^2/Z_0 is greater than, equal to, or less than one.

If we assume the $D^2/Z_0 < 1$, equation (20) is elliptic, and with a simple change of variables f is harmonic. The only harmonic functions with bounded derivatives everywhere are x, y, and a constant. Consequently, there are no solutions except these trivial ones, which correspond only to a uniform flow.

If $1 = D^2/Z_0$,

$$f_{yy} = 0 \qquad (21)$$

The solution to (21) is

$$f = a(x)y + b(x) \qquad (22)$$

We must consider the next order of approximation to determine f.

The terms in equations (3) and (4) proportional to σ^3 are

$$Dh_x + Z_2$$
$$= \tfrac{1}{2}(\text{Const.} + DZ_0^2\nabla_2^2 f_x - f_x^2 - f_y^2) \qquad (23)$$

and

$$DZ_{2x} + Z_0\nabla_2^2 h = (Z_0^3/3!)\nabla_2^4 f$$
$$- Z_1\nabla_2^2 f - Z_{1x}f_x - f_yZ_{1y} \qquad (24)$$

If we eliminate Z_2, we obtain

$$h_{yy} = -\frac{D}{Z_0}\left[\frac{DZ_0^2}{2}\nabla_2^2 f_{xx} - f_xf_{xx} - f_yf_{xy}\right]$$
$$+ \frac{1}{Z_0}\left[\frac{Z_0^3}{3!}\nabla_2^4 f + Z_1\nabla_2^2 f\right.$$
$$\left. - Z_{1x}f_x - Z_{1y}f_y\right] \qquad (25)$$

From (22) we find[3]

$$h_{yy} = -\frac{D}{Z_0}\left[\frac{DZ_0^2}{2}(a^{\mathrm{IV}}y + b^{\mathrm{IV}})\right.$$
$$\left. - (a^{\mathrm{I}}y + b^{\mathrm{I}})(a^{\mathrm{II}}y + b^{\mathrm{II}}) - aa^{\mathrm{I}}\right]$$
$$+ \frac{1}{Z_0}\left[\frac{Z_0^3}{3!}(a^{\mathrm{IV}}y + b^{\mathrm{IV}})\right.$$
$$+ D(a^{\mathrm{I}}y + b^{\mathrm{I}})(a^{\mathrm{II}}y + b^{\mathrm{II}})$$
$$- \mu D(a^{\mathrm{II}}y + b^{\mathrm{II}})$$
$$\left. + D(a^{\mathrm{II}}y + b^{\mathrm{II}})(a^{\mathrm{I}}y + b^{\mathrm{I}}) + Daa^{\mathrm{I}}\right] \qquad (26)$$

Z_1 has been eliminated by means of (18), and the constant in (18) has been put equal to μD.

Since h_y is bounded, we must have

$$h_{yy} = 0 \qquad (27a)$$

Then since the left-hand side of (26) is zero (and independent of y) we choose

$$a^{\mathrm{I}}(x) = 0 \qquad (27b)$$

We cannot put the coefficients of y and y^2 separately equal to zero, since there are two equations and only one unknown function, $a(x)$. Equation (26) simplifies to

$$D\frac{Z_0^2}{3}b^{\mathrm{IV}} - 3b^{\mathrm{I}}b^{\mathrm{II}} + \mu b^{\mathrm{II}} = 0 \qquad (28)$$

This equation may be integrated by quadratures with the aid of elliptic functions and $E(x)$, the elliptic integral of the second kind, as is indicated by Keller in his solution of the two-dimensional case.

We find, on integrating once, that

$$\frac{D}{3}Z_0^2 b^{\mathrm{III}} - \frac{3b^{I^2}}{2} + \mu b^{\mathrm{I}} + \frac{l}{2} = 0 \qquad (29)$$

where we have introduced a special form for the arbitrary constant in order to facilitate our later calculations. Multiplying by b^{II} and integrating again, we obtain

$$\frac{DZ_0^2 b^{\mathrm{II}^2}}{3} - b^{\mathrm{I}^3} + \mu b^{\mathrm{I}^2} + lb^{\mathrm{I}} - m = 0 \quad (30)$$

where we have again chosen a convenient form for the constant of integration. Solving for b^{II} we find

$$b^{\mathrm{II}}(x) = \frac{\sqrt{3}}{Z_0D^{\frac{1}{2}}}\sqrt{b^{\mathrm{I}^3} - \mu b^{\mathrm{I}^2} - lb^{\mathrm{I}} + m} \quad (31$$

or

$$b^{\mathrm{II}}(x) = L\sqrt{(r_1 - b^{\mathrm{I}})(r_2 - b^{\mathrm{I}})(b^{\mathrm{I}} - r_3)} \quad (32$$

where (note that $r_1 + r_2 + r_3 = -\mu$)

$$r_1 > r_2 > r_3 \qquad (33$$

and

$$L = \frac{\sqrt{3}}{Z_0D^{\frac{1}{2}}} \qquad (34$$

We see that b^{I} is an elliptic function. Since l

[3] We put

$$\frac{da(x)}{dx} = a^{\mathrm{I}}(x),\ \frac{d^2a(x)}{dx^2} = a^{\mathrm{II}}(x),\ \text{etc.}$$

; to be bounded, r_1, r_2, and r_3 must be real, nd r_3 negative. We find

$$b^{\mathrm{I}}(x) = r_2 - (r_2 - r_3)\, cn^2 Fx \qquad (35)$$

here

$$F = \frac{L(r_1 - r_3)^{\frac{1}{2}}}{2} \qquad (36)$$

he modulus k of the elliptic function is given by

$$k = \sqrt{(r_2 - r_3)/(r_1 - r_3)} \qquad (37)$$

The function $b(x)$ is then determined by an ıtegration,

$$b(x) = r_1 x - \frac{(r_2 - r_3)}{Fk^2} E(Fx, k) \qquad (38)$$

here

$$E(x) \equiv \int_0^x dn^2 x\, dx \qquad (39)$$

ıd we have chosen the constant of integration) that $b(0) = 0$.

These results are formally the same as those Keller for the two-dimensional solution. The ıly change in the final solution is the introuction into f and consequently into ϕ of the ɜrm $\sigma\ ay$. Since a is an arbitrary constant at ır disposal (of order of magnitude unity), we ay superimpose on the two-dimensional soluon an arbitrary current in the y direction of ıagnitude σ. It is clear from the arguments resented that the general solution, correct to l orders in σ, would have this same property, mply because of the conditions we have imposed ıat the derivative of the potential is bounded ꝛerywhere. We shall not repeat here the physiıl interpretation of the two-dimensional case hich is discussed in detail by Keller.

We have still to consider the most interesting ıse of equation (20); namely,

$$\frac{D^2}{Z_0} > 1 \qquad (40)$$

'e see that (20) is just the wave equation, which ıs infinitely many bounded solutions in two mensions. The general solution of (20) is, ı putting $\eta = \sqrt{(D^2 - Z_0)/Z_0}$,

$$f(x, y) = f^+(x + \eta y) + f^-(x - \eta y) \qquad (41)$$

as we know from the study of the two-dimensional wave equation. Very curious wave forms may be propagated, as, for example, a single ridge of arbitrary shape whose sections parallel to one of the characteristics of equation (20) are straight lines. (Fig. 2 gives one example.)

As in the two previous cases we have considered, the third approximation could be obtained from equations (23) and (24). We shall not do this, because the additional details would merely provide corrections to the second approximation and would not increase our understanding of the results. There seems to be no difficulty in principle about the calculation.

We indicate briefly the calculation of the adjustable parameters in the general case. The velocity potential is

$$\phi(x, y) = Dx + \sigma\{f^+(x + \eta y) + f^-(x - \eta y)\} + \cdots \qquad (42)$$

The equation of the free surface is $z - Z = 0$, where

$$Z(x, y) = Z_0 - \sigma D[f_x^{\ +}(x + \eta y) + f_x^{\ -}(x - \eta y)] + \cdots \qquad (43)$$

If λ_0 is the wavelength (for a periodic wave), the wave velocity C can be defined as the average value of the gradient of ϕ over one wavelength, at constant elevation. This definition agrees with that of *Stokes* [1905, p. 203]. We have

$$C = \frac{1}{\lambda_0} \int_0^{\lambda_0} \phi_x\, dx \qquad (44)$$

or, by integration,

$$C = \frac{1}{\lambda_0} [D\lambda_0 + \sigma\{f^+(\lambda_0 + \eta y) + f^-(\lambda_0 - \eta y) - f^+(\eta y) - f^-(\eta y)\} + \cdots] \qquad (45)$$

The wave height for a periodic wave is $Z_{\max} - Z_{\min}$,

$$H = Z_{\max} - Z_{\min} \qquad (46)$$

and the depth is the average value of Z over all the surface

$$d = \lim_{A \to \infty} \int_A \int dx\, dy\, Z(x, y) \qquad (47)$$

If the wave phenomenon is essentially aperiodic and in particular if it dies off at infinity, the wave velocity is the value of ϕ_x at infinity. The depth is also the value of Z at infinity. The wave height must be considered separately in each special case, as the example indicates. The wave velocity is the ratio of the wavelength to the period.

Examples (for the case $D^2/Z_0 > 1$)—We shall consider two examples which indicate the character of the results. First let us choose

$$f(x, y) = \cos \kappa x \cos \kappa\eta y \tag{48}$$

a special case of equation (46), for which we will construct the velocity potential and fix the adjustable parameters. From equation (24) we find

$$Z_1 = + D\kappa \sin \kappa x \cos \kappa\eta y \tag{49}$$

putting the constant of integration equal to zero. Thus we have

$$\phi = Dx + \sigma \cos \kappa x \cos \kappa\eta y \tag{50}$$

and

$$Z = Z_0 + \sigma D\kappa \sin \kappa x \cos \kappa\eta y \tag{51}$$

The surface is sketched in Figure 1.

If we reintroduce our dimensions, we obtain

$$\phi = \lambda\sqrt{gd}\left(\frac{Dx}{\lambda} + \frac{d^2}{\lambda^2}\cos\frac{\kappa x}{\lambda}\cos\frac{\kappa\eta y}{\lambda}\right) \tag{52}$$

and

$$Z = d\left[Z_0 + \frac{d^2}{\lambda^2} D\kappa \sin\frac{\kappa x}{\lambda}\cos\frac{\kappa\eta y}{\lambda}\right] \tag{53}$$

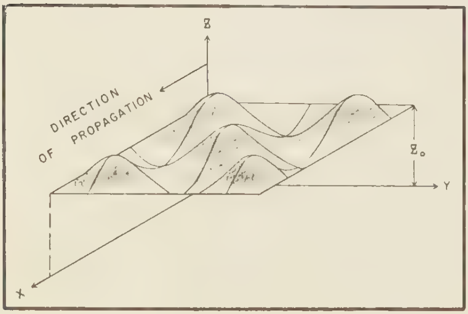


FIG. 1—A sketch of the ocean surface as given by $Z = Z_0 + \sigma\, D\kappa \sin \kappa x \cos \kappa\eta y$, illustrating the first example. Only the portion of the wave in first quadrant is shown.

Since the celerity is just the average value of the velocity in one wavelength,

$$C = \sqrt{gd}\, D \tag{54}$$

where C is the celerity. The wave height is

$$H = Z\left(+\frac{\pi\lambda}{2\kappa}, 0\right) - Z\left(-\frac{\pi\lambda}{2\kappa}, 0\right) \tag{55}$$

or

$$H = \frac{2d^3 D\kappa}{\lambda^2} \tag{56}$$

We have chosen here the wave height to be the maximum vertical distance from crest to trough. The depth is

$$d = \frac{1}{\left(\frac{2\pi\lambda}{\kappa}\right)^2 \frac{1}{\eta}} \int_0^{2\pi\lambda/\kappa} dx \cdot \int_0^{2\pi\lambda/\kappa\eta} dy\, Z(x, y) \tag{57}$$

On performing the integration, we obtain

$$d = dZ_0 \tag{58}$$

Thus we find

$$Z_0 = 1 \tag{59}$$

We see that the wavelength is

$$\lambda_0 = 2\pi\frac{\lambda}{\kappa} \qquad (60$$

We obtain from (54) the restriction that

$$\kappa < 2\pi \qquad (61$$

since physically our approximation was that th wavelength in the x direction was large in comparison with the depth. We see that in superposition of solutions of the form of equation (48 components of arbitrarily high wave numbe cannot be used. Thus the smallest scale of variation in any wave packet must be large in comparison with the depth. Finally we can obtai for the parameter D,

$$D = 2\frac{(2\pi)^2}{\kappa}\left(\frac{d}{H}\right)\left(\frac{d}{gT^2}\right) \qquad (6$$

and for λ_0, the wavelength,

$$\lambda_0 = 2\,\frac{(2\pi)^2}{\kappa}\,d\,\frac{d}{H}\sqrt{\frac{d}{gT^2}} \qquad (63)$$

by the simultaneous solution of equations (54) and (56), noting that λ_0/T is the celerity.

As a second example, we choose as f

$$f(x, y) = \int_0^\infty d\kappa e^{-\kappa} \cos \kappa(x + \eta y) \qquad (64)$$

or

$$f(x, y) = \frac{1}{1 + (x + \eta y)^2} \qquad (65)$$

which also has the form of equation (41) (with $f^- = 0$). Then we find

$$\phi(x, y) = Dx + \frac{d^2}{\lambda^2\{1 + (x + \eta y)^2\}} \qquad (66)$$

and

$$Z(x, y) = Z_0 + \frac{2d^2 D(x + \eta y)}{\lambda^2\{1 + (x + \eta y)^2\}^2} \qquad (67)$$

The surface is sketched in Figure 2. If the dimensions are reintroduced, we have

$$\phi(x, y) = \sqrt{gd}\left[Dx + \frac{d^2}{\lambda\{1 + (x + \eta y)^2\}}\right] \qquad (68)$$

and

$$Z(x, y) = d\left[Z_0 + \frac{2d^2 D\left(\frac{x}{\lambda} + \frac{\eta y}{\lambda}\right)}{\lambda^2\{1 + (x + \eta y)^2\}^2}\right] \qquad (69)$$

The celerity is ϕ_x for $x + \eta y$ large, or just

$$C = \sqrt{gd}\;D = \frac{\lambda_0}{T} \qquad (70)$$

The height is $Z_{\max} + Z_{\min}$, which is twice $Z_{\max}$. The maximum is at $(x + \eta y)/\lambda = +1/\sqrt{3}$, so that the height is

$$H = \frac{3\sqrt{3}}{4}\,d\,D\,\frac{d^2}{\lambda^2} \qquad (71)$$

Since the average of Z is dZ_0, we see that $Z_0 = 1$. If we call the wavelength twice the distance from the crest to the trough, we have

$$\lambda_0 = \tfrac{2}{3}\sqrt{3}\,\lambda \qquad (72)$$

Solving for λ_0 and D as a function of d, T, and H, we obtain

$$\lambda_0 = \sqrt{3}\,d\sqrt{\frac{d}{gT^2}}\,\frac{d}{H} \qquad (73)$$

and

$$D = \sqrt{3}\,\frac{d}{H}\,\frac{d}{gT^2} \qquad (74)$$

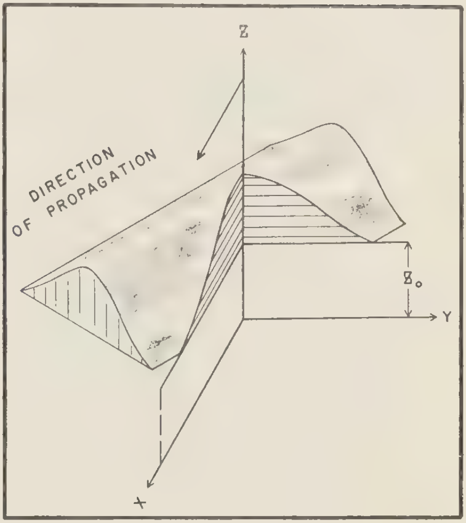


FIG. 2—A sketch of the ocean surface as given by $Z_0 + 2\dfrac{\sigma D(x + \eta y)}{[1 + (x + \eta y)^2]^2}$, illustrating the second example.

Concluding remarks—We have still to discuss the transition to a stationary coordinate system. If we let x', y', z' be the stationary coordinates, the transformation is

$$x' = x - Ct$$
$$y' = y \qquad (75)$$
$$z' = z$$

We see also that the x component of the velocity must be diminished by C; that is, the velocity potential is diminished by Cx.

The result in the two examples that $Z_0 = 1$ is not necessarily a general property of this problem. It depends on all the functions in $Z(x, y)$ having an average value of zero over one wavelength. Since we have considerable liberty in the choice of functions, no general conclusion is possible.

We have avoided the mathematical difficulty of discussing the convergence of the series expansion employed here. The convergence proofs for the two-dimensional case [*Friedrichs and Hyers* 1954; *Littman* 1957] depend in a decisive way on complex variable theory and do not admit a ready generalization in three dimensions. Some more general mathematical tools would seem to be necessary to construct a proof in this case. Especially since very curious effects have been predicted here, it seems important to seek both experimental and mathematical justification of these purely formal results.

It would also be of interest to settle the questions raised by equation (15c). While it is physically reasonable to assume that the first approximation is a uniform flow, there is no apparent physical reason why this must be the case. The results obtained in this paper all seem physically reasonable, even if rather unexpected; any future solutions should also satisfy this same test.

References

Stoker, J. J., *Water Waves,* Interscience Publishers, New York, 1957.

Keller, J. B., The solitary wave and periodic waves in shallow water, *Comm. on Pure and Applied Math., 1,* 323–339, 1948.

Friedrichs, K. O., and D. H. Hyers, The existence of the solitary wave, *Comm. on Pure and Applied Math., 7,* 517–550, 1954.

Littman, W., The existence of periodic waves near critical speed, *Comm. on Pure and Applied Math., 10,* 241–271, 1957.

Stokes, G. G., *Mathematical and Physical Papers,* Cambridge Univ. Press, *1,* 1905.

(Manuscript received June 19, 1959; presented at the Fourtieth Annual Meeting, Washington, D. C., May 5, 1959.)

Journal of Geophysical Research Volume 64, No. 11 November, 1959

Ice Petrofabric Observations from Blue Glacier, Washington, in Relation to Theory and Experiment*

W. Barclay Kamb

California Institute of Technology
Pasadena, California

Abstract—Ice observed on the surface of Blue Glacier is classified texturally into three types: *coarse bubbly ice, coarse clear ice,* and *fine ice.* The three types occur intercalated to form the observed foliated structure of the bulk glacier ice. Petrofabric study of fine ice reveals consistently a broad maximum in the density of *c*-axis orientations, centered about the pole of the foliation plane. This single-maximum fabric is in some respects similar to fabrics of stressed ice from polar glaciers, and the textures of fine ice and polar ice are similar. The fine-ice layers also resemble layers that have recently been produced by rapid shearing deformation of glacier ice in the laboratory. It is inferred that the fine-ice layers in the glacier constitute zones that are undergoing (or have recently undergone) rapid mechanical plastic flow, and that the adjacent coarse-ice layers originate by recrystallization from fine ice and are not now deforming rapidly by mechanical plastic flow (basal glide). Whether the fine-ice layers have predominately a tectonic origin or whether they originate predominately as infillings of snow in crevasses in the icefall is not known for certain.

Coarse bubbly ice fabrics generally show more than one maximum in the density of *c*-axis orientations. The statistical significance of multiple-maximum fabrics is tested by a comparison of several independent fabrics from given stress situations, and it is shown that the basic four-maximum pattern is reproducible, though subject to unexplained fluctuations in orientation. The 'diamond-shaped' four-maximum pattern is characteristic of ice subjected to long-continued shear stress of persistent orientation, and the long axis of the 'diamond' is (approximately) parallel to the direction of the stress vector that acts across the persistent plane of maximum shear stress. It is inferred that the basic features of the pattern develop at some depth within the glacier, and that subsequent deformation has affected the pattern to some extent.

The results of recent experimental studies of the origin of ice fabrics are in moderately good agreement with the Blue Glacier observations. Recent theoretical treatments are in sufficient disagreement to be ruled out.

A new method of presenting orientation data allows statistical inferences to be drawn directly from the fabric diagrams.

Introduction

The original studies by *Bader* [1951] and *Rigsby* [1951] of crystal orientation in glacier ce established the existence of the peculiar nultiple-maximum fabrics which, because they lefied reasonable explanation, have hampered n interpretive application of petrofabric techiques to glaciers. Later work [*Schwarzacher nd Untersteiner,* 1953; *Rigsby,* 1953; *Meier, Rigsby and Sharp,* 1954] added further supportng data, as well as further complications, and resented several hypotheses to explain the obervations, but the explanations did not seem fully satisfactory. The present investigation was undertaken with the original purpose of attempting to test critically whether the peculiar fabrics represent a real phenomenon or whether they may be due to the methods of observation and sampling. Two developments considerably broadened the scope and results of the work. First, the study described below of fine-grained ice has revealed a type of *c*-axis fabric not previously observed in temperate glacier ice but similar to *c*-axis fabrics reported by *Rigsby* [1955] from polar glaciers. Second, the progress of experimental studies of the origin of ice fabrics [*Steinemann,* 1958a, 1958b; *Rigsby,* 1958; *Shoumsky,* 1958] and the development of theoretical tools for analysis of the origin of

*Contribution No. 947, Division of the Geological Sciences, California Institute of Technology.

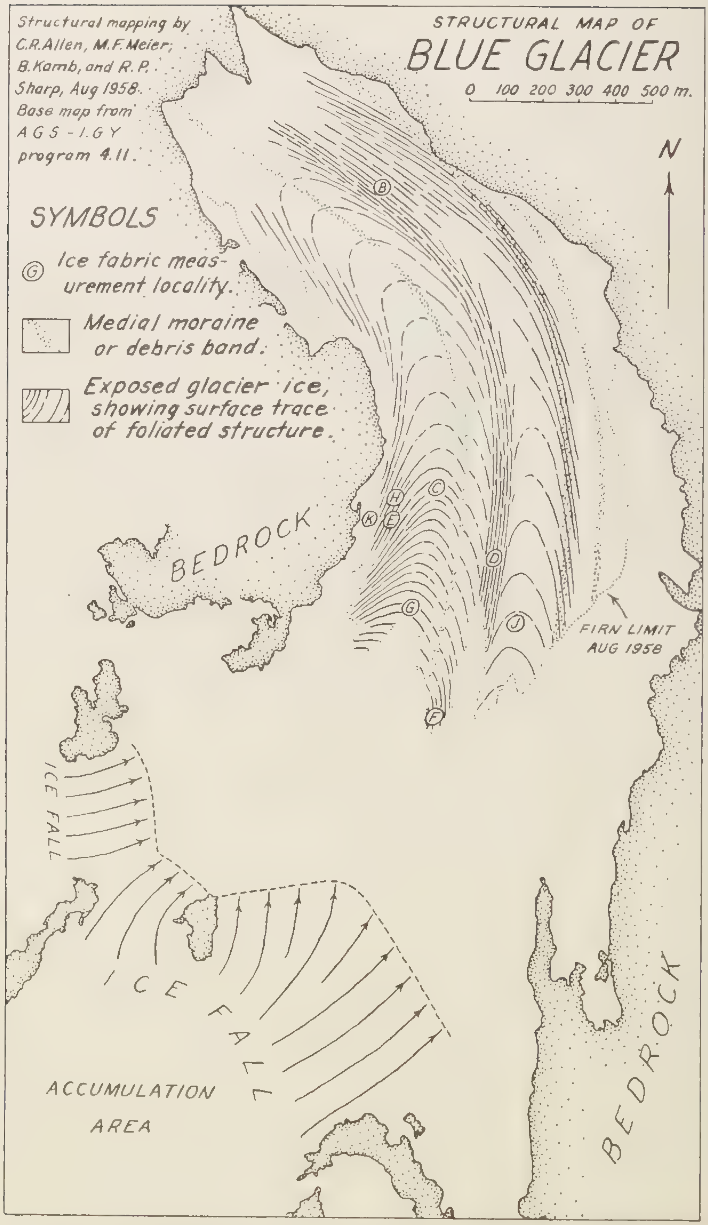


Fig. 1—Structural map of Blue Glacier, showing location of ice petrofabric measurements. The foliat structure dips nearly vertically (75° to 90°) at all the petrofabric localities.

preferred orientation [*Kamb*, 1959a; *MacDonald*, 1959; *Brace* (to be published)] has led to theoretical and empirical predictions to which critical observational tests could be applied. This paper presents the data obtained from Blue Glacier bearing on these problems, and reviews the current theoretical situation in the light of observational data.

Blue Glacier

Blue Glacier is located on Mount Olympus in Olympic National Park, Washington. The glacier originates on the upper slopes of the mountain in two separate accumulation basins, from which three separate ice streams descend in a precipitous ice fall to the valley below. There the ice streams unite to form a valley glacier 1 km wide, which flows down the valley to a terminus about 3 km from the base of the ice fall (Fig. 1). The present study concerns only the lower portion of the glacier, below the firn line, which in late summer lies about 1 km downstream from the base of the ice fall.

Textural Types of Glacier Ice

The solid ice exposed below the firn line may be classified texturally into three types, as follows.

(1) *Coarse bubbly ice:* coarse crystalline—crystals 1 to 6 cm in diameter as seen in cross section. Individual crystals have very elongate, sinuous or branching shapes and may extend 20 cm or more through the ice, traversing a complex path [*Bader*, 1951, p. 525]. The interdigitation of crystals of complex shapes results in a texture which does not disaggregate upon the loosening of the individual crystals by melting along the grain boundaries, as occurs on the 'weathered' glacier surface under the action of sunlight. Within the ice, and mostly within the individual crystals rather than along grain boundaries, are numerous spheroidal bubbles, 1 to 2 mm in diameter, which contain air and water vapor. The bubbles tend to cluster into indistinct layers ('bands'), giving rise to a weakly foliated structure. Sharply defined bubbly layers, 3 to 10 cm thick, occur relatively uncommonly; they contain a much greater bubble concentration than that of the normal coarse bubbly ice. The bubbles in these layers are finer, and the layers have a distinctive milk-white, almost porcelain-like appearance on the 'weathered' surface. The crystalline texture is completely uninfluenced by the bubble-layer foliation, the individual crystals extending through the bubble-rich and bubble-poor zones without any visible discontinuity. Coarse bubbly ice appears to be the most abundant type of ice found in temperate glaciers.

(2) *Coarse clear ice:* very coarse crystalline—crystals 3 to 12 cm in diameter; nearly bubble free, forming a clear, transparent aggregate; interlocking grain texture similar to that of the coarse bubbly ice. Coarse clear ice occurs in three structurally distinct situations in the glacier: (a) large zones of stagnant ice at the margins of the glacier; (b) thin (1 to 6 cm) layers intercalated with coarse bubbly ice in the foliated structure of the glacier, the coarse clear ice layers forming 'blue bands'[1]; (c) irregular pods and masses contained within fine ice, usually elongated parallel to the foliation; less commonly as irregular pods within coarse bubbly ice. The coarse clear ice weathers to a clear, irregular surface marked by a pattern of Forel stripes [*von Klebelsberg*, 1948, p. 41] on the individual crystals.

(3) *Fine ice:* crystals 0.5 to 2 mm in diameter, rarely up to 4 mm in maximum dimension; individual grains equant to somewhat discoid in shape; the crystals do not interdigitate or interlock, so that they disaggregate upon 'weathering'; contains a scattering of small bubbles less than 1 mm in size, showing weak to strong segregation into bubble-rich and bubble-poor layers parallel to the foliation. The 'weathered' fine ice is a granular mass identical in appearance to firn; the designation *firneis* would be appropriate except for the necessarily implied genetic relation to actual firn; also, unlike firn, the unweathered fine ice is a clear, dense, impermeable solid. Fine ice occurs as layers generally 2 to 15 cm thick intercalated with coarse bubbly ice and coarse clear ice to form a foliated structure, or foliation, which can be observed almost everywhere on the glacier surface. The fine-ice layers are probably equivalent to similar layers that

[1] The term 'blue band' is also applied to bubble-poor zones within the coarse bubbly ice masses, and because of this ambiguity I have avoided using the term.

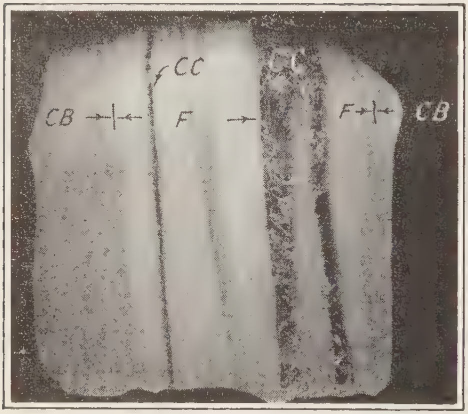


Fig. 2—Slab of glacier ice, about 30 cm on edge, showing coarse bubbly ice (CB), coarse clear ice (CC), and fine ice (F). Locality *E*. Photograph courtesy of C. R. Allen.

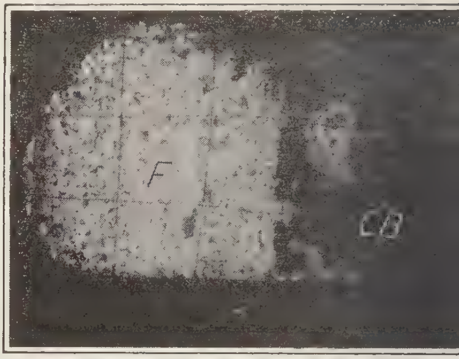


Fig. 3—Thin-section of a portion of the ice sla of Figure 3, showing the texture of fine ice (F and coarse bubbly ice (CB). The grid lines ar 1 inch apart. Crossed nicols.

have been called 'granulated ice' [*Rigsby*, 1953, p. 27] or 'brecciated ice' [*Meier*, 1957, p. 116; also personal communication] in other glaciers. However, the fine ice shows no manifest indication, in thin section, of brecciation or cataclasis. Individual crystals in most samples of fine ice tend to be dimensionally shortened in the direction perpendicular to the foliation plane.

In Figure 2 a typical specimen of glacier ice is shown, illustrating the three types of ice and the foliated structure that they compose. The types of ice are compared in thin section in Figure 3, and in Figure 4 the detailed textural features of the fine ice are illustrated.

Procedure of the Fabric Study

Ice *c*-axis fabrics have been measured at nine localities on Blue Glacier, indicated by the lettered circles in Figure 1. At some of the localities several separate fabrics were measured. In particular, the fabrics of fine ice and adjacent coarse bubbly ice were measured for comparison at several localities. In the present paper only fabrics of fine ice and coarse bubbly ice are considered, with emphasis on the relation of these fabrics to the states of stress under which they develop and to general questions of the origin of ice fabrics. The relation between ice fabric and structure in the glacier as a whole is treated in a separate paper [*Allen, Kamb, Meier, and Sharp* to be published] in which the strikin foliation pattern displayed in Figure 1 is ana lyzed. The fabric data obtained from Blu Glacier are summarized in Table 1, and th particular data discussed in the present pape are indicated there.

Orientations of ice *c*-axes were measured b the universal stage technique described b *Bader* [1951] and *Rigsby* [1953]. The *c*-ax orientations are plotted on the lower hemisphe of an azimuthal equal-area (Schmidt) proje tion; each diagram is identified by fabric numb (Table 1) and by locality, and the number c-axes measured and the number of thin sectio required for the measurements are also give Rather than rotating the many diagrams fro the projection planes in which they were mea ured into the horizontal projection plane, standard system of designating the orientatio of the projections is used for presentation of t data. At each locality a local coordinate sy tem, consisting of orthogonal axes labeled *n* *e-w*, and *up-dn* is set up according to the f lowing specifications: (1) Two of the axes a chosen to be in the (always nearly vertic foliation plane, the third normal thereto. (One of the axes in the foliation plane is tak horizontal (strike line), and this axis is label either *n-s* or *e-w* according as the strike of t foliation is more nearly parallel to the leng of the glacier (*n-s*, *n* being taken downglaci or transverse to it (*e-w*, *w* being taken towa

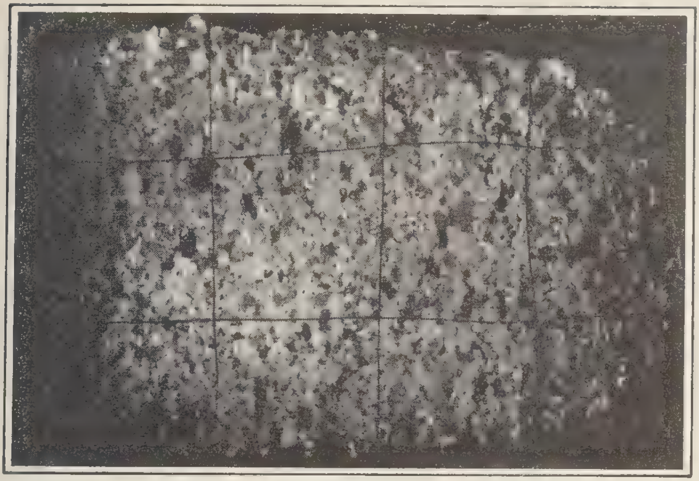


Fig. 4—Thin-section of fine ice from locality *F*. Grid lines are 1 inch apart, on the average. Crossed ols. Note the preferential flattening of the grains perpendicular to the foliation plane (which is perndicular to the thin-section, its trace vertical in the photograph).

e left bank). (3) The (always nearly horizonl) axis perpendicular to the foliation plane is rrespondingly labeled respectively *e-w* or *n-s*, d the third (nearly vertical) axis is labeled -*dn*.

In all cases the data are presented as scatter grams showing the individual *c*-axis orientans that were measured, and in some cases the rresponding contoured density diagrams are own. The scatter diagrams are given because a statistical examination of the data the scat diagrams should be used rather than conured density diagrams; also, the scatter diams give a more informative picture of the ual character of the orientation distribution, d they lend themselves to a further advangeous refinement to be mentioned later. The nsity diagrams are prepared by a somewhat vel procedure, described in the Appendix, h that statistical inferences can be drawn ectly from the diagrams. The contoured diams are given only in cases where a statistical mination seems to be called for.

The measured *c*-axis inclinations (to the plane each thin section) were corrected for index refraction by Snell's law, before plotting. gsby [reported by *Langway*, 1958, p. 7] has nd experimentally that the Snell's law cortion is not applicable to orientations measured by setting the *c*-axis horizontal, and recently the complete theoretical explanation of this phenomenon has been found [*Kamb*, 1959b]. At the time the fabric diagrams were prepared, the proper correction procedure was not known. A replotting of the diagrams in the light of the correct procedure, though desirable, is not considered essential for the qualitative and semiquantitative purposes to which they are applied and was therefore not undertaken. Nevertheless, it is emphasized that for a precise quantitative study of the angular relations exhibited in the diagrams the original data should be replotted.

According to theory [*Kamb*, 1959b], a gap about 10° in width should be found in the fabric diagrams, as here plotted, at about 45° to the pole of the plane of the thin section, a feature first noticed in practice by *Rigsby* [1951, p. 597–598]. In the present work such a gap does not occur, because it was found in practice that poles lying in the vicinity of the gap can be measured both by setting *c* horizontal and by setting it vertical, and, because the resulting corrected inclinations were averaged, the discrepancy in the neighborhood of the gap region is blurred out. The maximum error of plotting is about 5°. Some diagrams represent combined measurements from thin sections of different

TABLE 1—*Fabric data from Blue Glacier*

Locality	Fabric number	Type of ice	Number of c-axes	Number of thin sect.	Type of fabric	Figure number
B	C-1	coarse bubbly	239	18	4-max.	13
C	C-2	coarse bubbly	240	12	multiple max.	...
D	C-3	coarse bubbly	134	11	4-max.	9
	C-4	coarse bubbly	122	7	4-max.	9
	F-1-2	fine	335	4	single broad max.	5
E	S-1	coarse bubbly	125	10	4-max.	11
	F-3	fine	296	12	single broad max.	5
F	F-4	fine	149	6	single broad max.	5
	C-6	coarse bubbly	120	4	diffuse 4 (?)-max.	10
	C-7	coarse bubbly	100	4	4-max.	10
G	C-8	coarse bubbly (intra-ogive)	229	7	single broad max.	...
	C-9	coarse bubbly (inter-ogive)	421	19	multiple max.	...
	F-5	fine	210	5	single broad max.	7
H	S-2	coarse bubbly	106	4	4-max.	11
	S-3	coarse bubbly	75	4	4-max.	11
	S-4	coarse bubbly	59	3	4 (?)-max.	11
	S-5	coarse bubbly	97	5	4-max.	11
J	C-14-15	coarse bubbly	149	7	single broad irregular max.	...
K	F-6	fine	125	2	single broad max.	6

orientation, and this tends further to obliterate any suggestion of a gap at 45° to the pole of the final projection.

FINE-ICE FABRIC

Because of its relatively small grain size, the fine ice is difficult to study by standard ice petrofabric technique. Perhaps this is why ice of similar type in other glaciers does not appear to have been studied previously. The principal difficulty is that a certain percentage of the grains (perhaps up to 25% by volume of the ice) are too small to measure except in the very thinnest parts of the thin section, where the i soon disaggregates. Even for relatively large to 4 mm) crystals, the attempt to measure o entations accurately without magnification taxing, and the accuracy of measurement doubtless poorer than for the large crystals the coarse ice. To avoid possible influence of t orientation of the thin section on the type fabric pattern obtained, I have in several stances measured fabrics both from thin s tions cut parallel and from sections cut p pendicular to the foliation, from the same sample. The patterns so obtained do not di

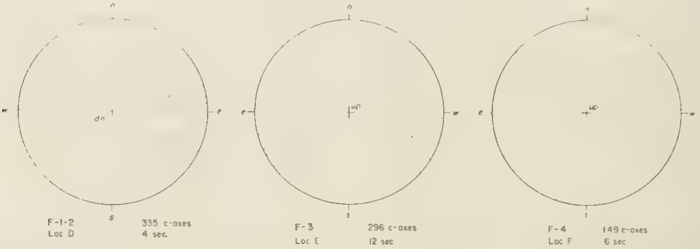


FIG. 5 -Fabrics of fine ice from localities *D, E,* and *F*. Orientation of coordinate axes is describe the text. Pole of the foliation plane is in each case *w* (or *e*)

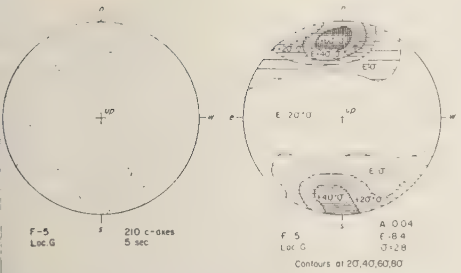


Fig. 6—Scatter diagram and contoured density iagram for fine ice from locality *G*. Contouring rocedure and interpretation are described in the ppendix. The discrepancy between the center of reatest orientation-density and the pole of the oliation plane, which is at *n* (or *s*), is not sigificant statistically.

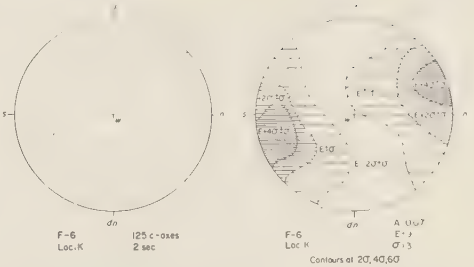


Fig. 7—Ice *c*-axis orientations in fine ice from a crevasse filling, locality *K*. The pole of the crevasse wall is at *s* (or *n*). The sense and orientation of shear, indicated by drag, is such that the *s* wall of the crevasse has moved *w* with respect to the *n* wall.

gnificantly, in spite of the obvious possible ource of difference (the much greater ease and ccuracy of measurement for crystals whose -axes are inclined less than about 45° to the ormal to the thin section than for crystals of reater inclination).

The observed fine-ice fabrics are remarkably mple and consistent. Each of the fine-ice ıbrics measured (Figs. 5 to 7) shows the same asic pattern, a broad maximum in the density f *c*-axes orientations, centered, within the staıstical uncertainty, about the pole of the foliaıon plane, regardless of the attitude of the oliation or its structural position in the glacier. he preferred orientation becomes well defined ı scatter diagrams that contain more than 200 easured *c*-axes, but for fewer axes the prerred orientation is less striking. To test the egree of preferred orientation in the fabric ith the fewest *c*-axes (125), a contoured denty diagram is given in Figure 7; for comparin, a density diagram based on 210 *c*-axes is so given (Figure 6). In Figure 7 the *c*-axis aximum constitutes a zone in which the denty of *c*-axes is systematically greater by one five standard deviations than the density pected for a random sample of 125 *c*-axes om a population having no preferred orientaon; and likewise the girdle of minimum *c*-axis nsity represents densities systematically low one to three standard deviations. In Figure 6 e preferred orientation is perhaps even more onounced, though the differences between the density diagrams in Figures 6 and 7 may be due mainly to the greater resolution of the latter, owing to the larger number of *c*-axes and its effect on the contouring procedure, as discussed in the Appendix.

The reproducibility of the single broad maximum in *c*-axis orientation density within fine ice from the various structural situations in the glacier strongly suggests that this pattern is a universal feature of the fine ice and may be expected from any fine-ice layer in the glacier. Such a pattern has not, apparently, been observed previously in temperate glacier ice.

From a polar glacier on Greenland, *Rigsby* [1955] reported a fabric ("Nuna loc. 1") in which the ice *c*-axes were strongly clustered in a single narrow maximum about the pole to the foliation plane, and recently he has reported an even stronger single maximum in ice from a tunnel in the Greenland ice sheet [*Rigsby*, 1958, p. 354]. The polar ice is relatively fine-grained, individual crystals being 5 mm or less in diameter, so that thin sections illustrated by *Rigsby* [1955, Figures 5 and 6] resemble sections of fine ice from Blue Glacier.

Although the fabrics of both types of ice (polar ice and fine temperate ice) consist of a single maximum of *c*-axis orientations centered about the pole of the foliation plane, in polar ice the maximum is sharp and narrow, and it contains *c*-axis densities of over 30% per 1% of area, whereas in fine temperate ice the maximum is characteristically very broad and uniform, with densities not exceeding 5% per 1%

of area. In addition, fine ice appears to contain a greater amount of the (relatively) finest grains (less than 1 mm in size) than polar ice contains.

The individual crystals of fine ice invariably show some degree of dimensional orientation. The grains are flattened perpendicular to the foliation plane—sometimes strikingly so, as shown in Figure 4. The dimensional orientation is reflected by a subtle banding visible on 'weathered' surfaces of fine ice. These observations appear to be the first recognition of well-defined dimensional orientation in glacier ice [*Rigsby*, 1958, p. 357].

The relation between the fine-ice layers and adjacent layers of coarse bubbly ice is a suggestive one. The fabric of the fine ice is just what would be expected to result from mechanical plastic flow, due to basal glide, in simple shear across the foliation plane. The multiple-maximum fabric of the adjacent coarse ice, containing no ice crystals oriented in the most favorable way for basal glide, hardly seems explicable on the basis of the mechanical plastic flow process. Consequently it is tempting to conclude that the layers of fine ice are layers in which the ice has been deforming mainly by active mechanical plastic flow, and that recrystallization of the fine ice leads to coarse bubbly ice, which, because of the unfavorable orientation of the individual ice crystals that it contains, no longer deforms rapidly by mechanical plastic flow.

This conclusion is in harmony with the observation that fine-ice layers are more abundant and thicker in the marginal, most-actively-deforming portions of the glacier than toward the center, where the rate of deformation (near the surface) is small. An exception is the 'foliation septum' (the zone of longitudinal foliation that passes through locality *D*, Figure 1), which, toward the head of the glacier, contains abundant thick layers of fine ice. However, these are probably inherited from the process by which the foliation septum is produced; they disappear down-glacier. Toward the snout of the glacier, fine-ice layers are essentially absent near the center line but become abundant toward the margins.

The relation between ice texture and rate of deformation has been noted by *Rigsby* [1958, p. 357], who summarizes his laboratory experience and extensive field observations as follows 'Crystal size increases with temperature and time in glaciers in the inactive or stagnant state whereas the size decreases with increasing strain rates of the more active glaciers. Laboratory experiments show that deformation of polycrystalline ice tends to decrease the grain size but melting temperatures tend to rapidly recrystallize this ice which has been deformed into larger crystals.' *Shoumsky* [1958, p. 246] has illustrated experimentally in a striking manner the reduction in crystal size that takes place within 'shear zones' developed in ice specimens loaded tangentially. The 'shear zones' in his experimentally deformed specimens show a striking textural similarity to fine-ice layers in Blue Glacier, and the *c*-axis fabrics appear to be very similar also.

The above ideas imply a difference in the mechanical properties of coarse ice and fine ice It seems also that each type of ice should have a characteristic anisotropy to plastic deformation. Such differences and anisotropies have no been demonstrated, although the marked plasti anisotropy of single ice crystals [*Steinemann* 1954; *Nakaya*, 1958] should doubtless be reflected in anisotropy of polycrystalline aggregates that have preferred orientation.

Also implied is a marked inhomogeneity, on scale of the order of centimenters, in the rate o deformation of ice in the glacier. Such an inhomogeneity of flow has not been noticec though the subject is under investigation o other glaciers (N. Untersteiner, personal communication). On Blue Glacier, no structura features that are transverse to the foliatior except crevasses, have been found which coul give evidence of small-scale non-uniformities i the velocity profile. Crevasses show no indication of offset along fine-ice layers, so that if nonuniformities in flow exist they must be confine to the bulk of the glacier in which the stre pattern is not modified by crevassing. On Sa katchewan Glacier, however, *Meier* [1957, 117] has described and illustrated wrinklir and offsetting of primary structures by nonuniform shearing motions across planes parall to foliation, although he did not study this effe in relation to the presence or position of fin ice layers.

It is interesting that *Rigsby* [1955], as a reult of his study of ice fabrics in Greenland aciers, has inferred a relation between polar e (sub-freezing-point ice) and temperate ice hat is analogous to the relation in Blue Glacier etween fine ice and coarse bubbly ice. The ultiple-maximum fabrics of temperate ice near he Greenland coast originate, he infers, by crystallization from the single-maximum fabc of stressed polar ice. Nevertheless, significant stinctions between polar ice and fine (tempere) ice exist. In addition to the fabric charcter, discussed above, there is the reported *Rigsby,* 1958, p. 357] frequency of strain hadows in polar ice. Strain shadows are essenally absent in fine ice. Since plastic deformaon without annealing recrystallization must evitably lead to strain shadows in the polyystalline aggregate, it appears that annealing crystallization, or recrystallization of some pe, must in any case play a part in the dermation of fine ice *in situ*.

The possibility should be considered that the ne-ice layers originate not directly in the flow rocess of the glacier, as surmised above, but stead as infillings of snow in crevasses. This ossibility receives concrete support from the bservation of fine-ice layers that recognizably ave such an origin. The geometry of these evasse fillings is shown in Figure 8. The fillings cur only in the outer part of the zone of arginal crevassing, where the foliation pattern thoroughly disrupted by faulting and folding, nd where the ice is predominantly the coarse ear type (stagnant or near stagnant). The llings occupy crevasses which have been roted 40° or more from their original trend, and ave been closed, in the course of the over-all ear deformation at the glacier margin. A few llings have been rotated until nearly parallel the glacier margin.

It is clear that such fillings do not account r the bulk of the fine ice in the glacier: first, ecause the actual number of fillings is small en in the crevassed zone, where the ice is aproaching stagnancy; and second, because the rge amount (about 20% by volume) of fine e in the actively deforming zone centerward om the crevassed zone has never been affected marginal crevassing. The possibility remains at the fine ice originates, at least in part, as crevasse fillings in the ice fall. Such an origin may explain the relative abundance of fine ice in the ogives, (which are part of the structural pattern seen in Fig. 1), by appeal to an annual process. It does not explain the greater abundance of fine-ice layers toward the glacier margin, all the way to the terminus. It also cannot explain the preservation of unrecrystallized fine ice under the prevailing temperature conditions, which must favor rapid increase of crystal size [*Perutz and Seligman,* 1939] in the absence of the disruptive effect of rapid plastic flow. Consequently the tectonic role of the fine-ice layers, as suggested above, must be called upon, it appears, regardless of the ultimate origin of the layers themselves.[2]

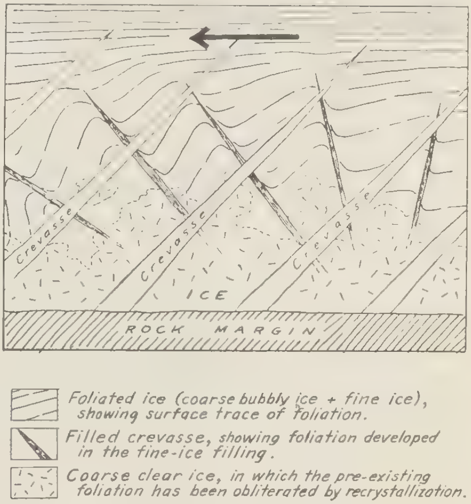

crevasse
crevasse
crevasse
ICE
ROCK MARGIN
Foliated ice (coarse bubbly ice + fine ice), showing surface trace of foliation.
Filled crevasse, showing foliation developed in the fine-ice filling.
Coarse clear ice, in which the pre-existing foliation has been obliterated by recrystallization.

Fig. 8—Schematic diagram showing, in plan view, the marginal portion of Blue Glacier in the vicinity of locality *K*. The direction of ice motion is indicated by the arrow. Open crevasses are labeled, and the width of the crevassed zone is about 100 meters. The foliation dips nearly vertically everywhere in the diagram.

The crevasse fillings of fine ice provide opportunity to observe the type of *c*-axis fabric produced naturally under a well-defined combination of shear stress and normal stress. The shear

[2] Preliminary results of a study, now under way [*Epstein and Sharp,* 1959], of the oxygen-isotopic composition of the fine- and coarse-ice layers lend some support to the crevasse-filling hypothesis of the origin of fine ice.

deformation is indicated by drag, which is visible in the pre-existing foliation adjacent to the walls of the filled crevasse. The drag is a consistent feature of the filled crevasses, and the geometry of the drag pattern (Fig. 8) shows that the drag originates in the process of rotation of adjacent large blocks of ice bounded by marginal crevasses, the rotation being effected by the overall shearing motion of the uncrevassed ice below and by the downglacier motion of the ice centerward from the crevassed zone. Figure 7 shows the fabric measured upon a sample of fine ice from a crevasse filling, the orientation of the pattern relative to the applied stress being described in the caption. The fabric is, within statistical uncertainty, identical to fabrics of fine-ice layers incorporated in the foliated structure of uncrevassed ice.

Coarse-Ice Fabrics

At every locality at which fine-ice fabrics were measured (except *K*), the adjacent coarse bubbly ice was also studied petrofabrically. In general it is found that the coarse bubbly ice has a fabric containing several maxima in the density of *c*-axis orientations. Some of these multiple-maximum fabrics are of the 'ideal' four-maximum type found by *Rigsby* [1951, 1953], whereas others are unrelated to the four-maximum type. Ideal four-maximum fabrics are found consistently in ice that has been subjected to long-continued strain in simple shear. In the present paper, only fabrics of coarse bubbly ice of this type are considered.

Statistical Validity of Multiple-Maximum Fabrics

Concern as to the statistical validity of multiple-maximum fabrics arises from the peculiar texture of the coarse ice. Individual crystals are complex in shape and may extend a distance of 20 cm or farther through the polycrystalline aggregate [*Bader*, 1951, Figure 2]. Consequently, if a coarse-ice fabric is obtained from a specimen of convenient size (less than 30 cm on edge) there is a strong likelihood that the actual number of crystals sampled may be small, a given crystal being measured many times where it is intersected in the successive sections or at widely separated places in the same section. The resulting sampling situation could give rise to an apparent multiple-maximum type of fabric. Al though the sampling efficiency was improved b measuring orientations only in slabs cut 4 to inches apart laterally (*Rigsby* [1953, p. 4] use a 2-inch spacing), the sampling uncertainty ca only be overcome by other procedures, describe below.

Another sampling problem results from th great variation in size of individual crystals a seen in thin section. The individual crysta range in cross-sectional area from less tha 1 cm^2 to over 50 cm^2. An interpretation of th observed fabric patterns would logically de with the volumetric proportions of various or entations in the polycrystal, whereas standa petrofabric procedure leads to orientation dens ties in terms of the percentage of individua which have the various orientations. The tw quantities evidently may differ greatly if th variation in grain size in the aggregate is larg as would happen, for example, if due to favo able orientation in relation to applied stress t larger crystals had grown at the expense of t smaller. To provide for an analysis of the petr fabric results from either standpoint, I ha utilized the scatter diagrams to show not on the orientation of each crystal measured, b also the approximate areal extent of each, a cording to a system described in the captions the figures. (The few coarse-ice diagrams that not show the size distinctions were measured b fore this system had become standard practic

The statistical matter of most importance the degree of reproducibility of the multip maximum fabrics formed under given stress co ditions. This has been tested at four localities Blue Glacier. At localities *D* and *F*, fabrics ha been measured from adjacent samples of coa ice. The pairs of coarse-ice samples were tained from coarse-ice layers which, althou adjacent (within 3 meters of one another), w physically separated by one or more layers fine ice, so that the orientations in each pair samples are certainly independent. The res ing pairs of fabrics are shown in Figures 9 a 10.

The most comprehensive test of fabric producibility was carried out at localities *E* *H*, near the west margin of the glacier in zone of strongly developed marginal foliat The results are shown in Figure 11. The f

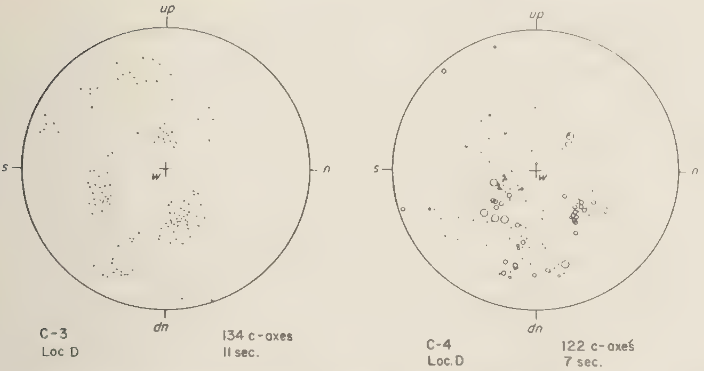


Fig. 9—Comparison of independent fabrics of coarse bubbly ice from locality *D*. In C-4, the cross-
ctional area α (in square cm) of each crystal, as seen in thin-section, is indicated according to the
llowing scheme: solid dots, $0.5 < \alpha < 3$; small open circles, $3 < \alpha < 6$; medium circles, $6 < \alpha < 15$;
rge circles, $15 < \alpha < 50$. w is the pole of the foliation plane.

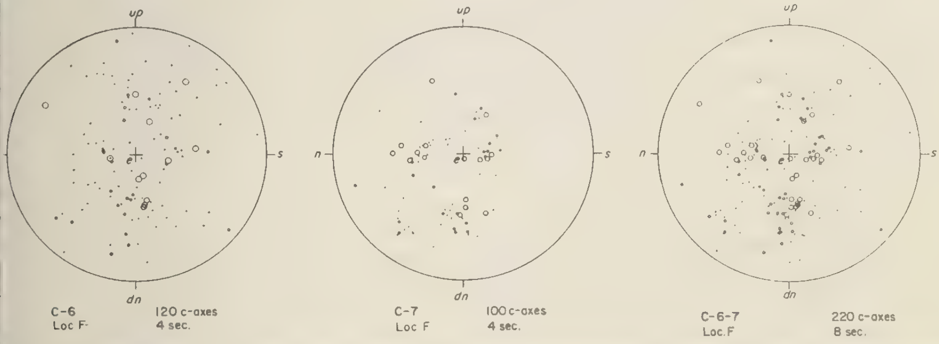


Fig. 10—Comparison of independent fabrics of coarse bubbly ice from locality *F*. C-6-7 is a composite
C-6 and C-7. Cross-sectional area α (in square cm) of individual crystals is shown as follows: solid
ots, $0.5 < \alpha < 3$; small open circles, $3 < \alpha < 15$; large circles $15 < \alpha < 50$. e is the pole of the
liation plane.

brics from locality *H* represent ice samples
parated from one another by distances of 5 to
) meters normal to the foliation plane, and
ence separated by many fine-ice layers. Lo-
lity *E* lies about 50 meters south of locality
; the stress situation at the two localities is,
r practical purposes, identical. Although the
ve diagrams in Figure 11 are by no means
entical, the degree of similarity is striking, and
e basic pattern is well defined. The similarity
etween the two fabrics of Figure 9 also is
rong, but between those of Figure 10 it is less
. In the latter case a multiple-maximum fabric
not well developed in fabric C-6, but the two fabrics C-6 and C-7 are sufficiently similar that, when combined (as shown), the four-maximum pattern of C-7 is actually somewhat enhanced.

The above comparisons establish the fact that multiple-maximum fabrics are a sufficiently reproducible feature of the stress situation under which they develop, sufficiently reproducible both with respect to the extent of development of the various maxima and to the geometry and orientation of the pattern as a whole, to be useful in deducing the stress situation under which they originate. A similar degree of reproducibility has been briefly reported by *Schwarzacher and Untersteiner* [1953, p. 117], although the

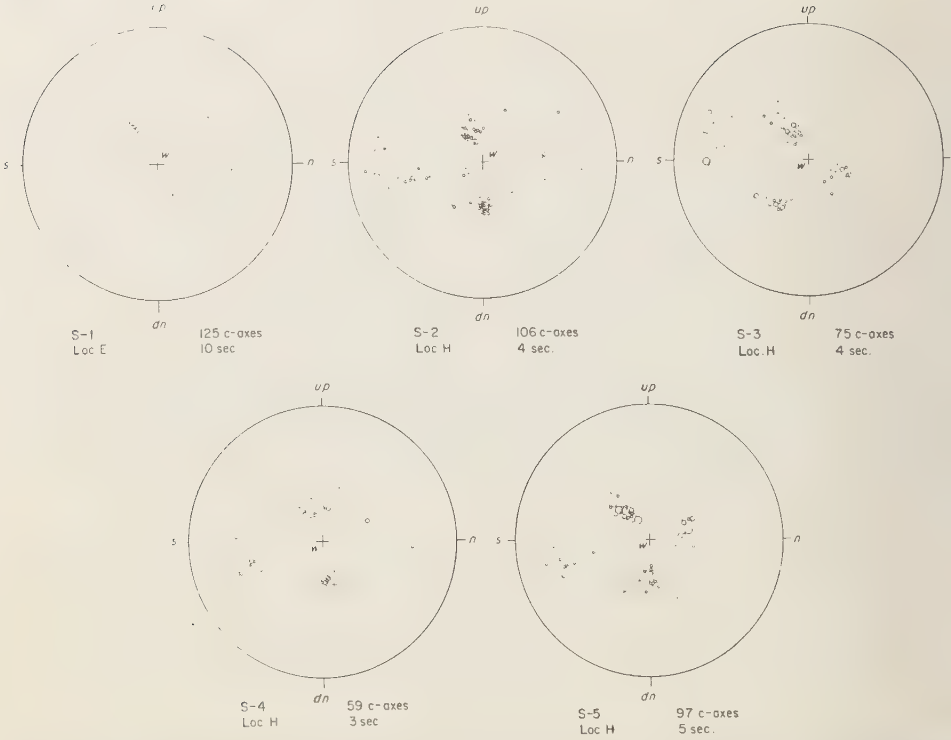


Fig. 11—Comparison of fabrics of coarse bubbly ice from the zone of strongly developed foliatic near the western margin of the glacier, at the outer edge of the zone of marginal crevassing (localiti *E* and *H*). Cross-sectional area α (in square cm) of each crystal is shown as follows: solid dots, 1 < $\alpha < 3$; small open circles, $3 < \alpha < 6$; intermediate circles, $6 < \alpha < 12$; large circles, $12 < \alpha < 3$ Areal designations omitted in S-1. w is the pole of the foliation plane, and the reference axes have tl same spatial orientation in all of the samples studied.

fabrics that they observed were quite different from the ideal four-maximum fabrics reported by *Rigsby* [1953].

The amount of variation in orientation of the multiple-maximum patterns obtained from nearby ice samples, as shown particularly in Figures 9 and 11, is considerably greater, however, than can be expected from the accuracy to which the orientation of the thin sections is known and from the accuracy of measurement of c-axis orientations. Significantly, the relative orientation of the several maxima in a given fabric, with respect to one another, is more consistent than the orientation of the patterns as a whole with respect to the foliation plane (which in this locality of well-developed, undisturbed foliation bears a constant orientation relative to the applied stress), confirming early observ. tions by *Rigsby* [1951, pp. 595–596]. It appea that the local control of the relative orientatic and degree of development of the several ma. ima is stronger than the over-all control due the stress situation and foliation pattern. Pe haps this effect is to be ascribed to locally var ing rotations (with respect to the foliatic plane) superimposed on the ice by the shearii motion that has taken place since the main fe tures of the fabric developed.

Relation Between Fabric Type and Stress Situation

To test theoretical and experimental a proaches to the problem of the origin of t coarse-ice fabrics, a comparison is needed heoretically or empirically predicted fabrics vith fabrics developed naturally under well-lefined stress conditions. An attempt to make uch a test from existing petrofabric data in the iterature does not prove possible, because most f the fabrics reported have been measured on amples of ice for which the precise stress situaion at the time of recrystallization, to the exent that it can be judged from criteria indeendent of the condition of the ice itself, is nknown. The only possible exceptions, to my nowledge, are certain of the fabric measurenents made by Rigsby on Saskatchewan Glacier *Meier, Rigsby, and Sharp,* 1954, pp. 22–23]. But unfortunately these potentially useful fabric atterns are anomalous: for reasons that can nly be surmised, the fabrics are weakly deeloped, they are aberrant from the ideal multile-maximum fabric type, and they fail to show consistent pattern.

The coarse bubbly ice fabrics at localities *B*, *E*, and *H* were obtained specifically to provide he necessary information. Of the accessible surace of Blue Glacier, the only portions for which he stress situation is well defined by available nformation are the rapidly shearing zones ear the lateral margins, at the inner edge of the zone of marginal crevassing. Localities *B, E,* and *H* occupy situations of this kind. Ice at the surface in the central part of the glacier has probably experienced a complex stress history, as indicated by the complexities of the foliation pattern. Toward the outer part of the zone of marginal crevassing, the ice has been much disturbed by faulting and folding, and its stress history is, again, complex.

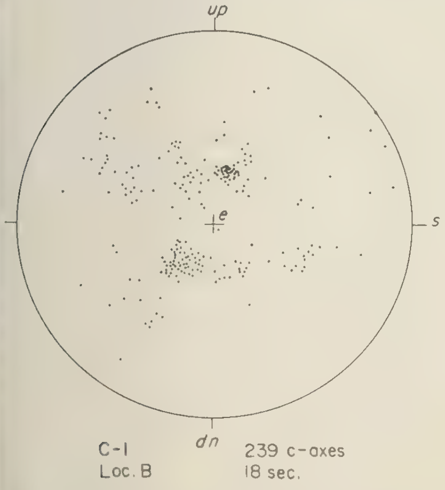


FIG. 12—Fabric of coarse bubbly ice from near he eastern margin of the glacier, about 50 meters enterward from the inner limit of the zone of arginal crevassing (locality *B*). Crystal sizes not hown. Pole of the foliation plane is *e*.

The consistent fabric patterns obtained at localities *E* and *H* (Fig. 11) indicate that the ideal four-maximum pattern develops under conditions of strong shear stress that maintains a persistent orientation. This is confirmed at locality *B*, which is on the opposite side of the glacier from *E* and *H* and farther down the glacier, where the same basic pattern is obtained (Fig. 12). When compared with fabrics obtained elsewhere on the glacier, fabrics which in many cases differ greatly from the pattern obtained at *B, E,* and *H,* these results reaffirm the conclusion advanced originally by *Rigsby* [1951, p. 595], that the diamond-shaped four-maximum patterns are characteristic of ice subjected to shear stress.

With the Blue Glacier data we can go beyond Rigsby's original conclusion and consider in detail the relation between fabric and applied shear stress. In Figures 11 and 12, the plane of each of the diagrams is the persistent plane of maximum shear stress, and the stress vector acting across this plane (from the side of the viewer) acts toward the axis *n*. In accordance with the stress symmetry, the fabrics should show a *s w n* (or *s e n*) mirror plane (monoclinic symmetry), and in fact they do, approximately. In general the (approximate) mirror plane seems to dip gently toward *s*. This perhaps reflects the fact that the velocity vector of the ice near the glacier margin has a significant upward component, as is found on other glaciers [*Meier,* 1957, p. 52], so that at depth the shear stress vector (which is oriented in accordance with the derivative, in the *e-w* direction, of the velocity vector) acts with an upward component across the *s-up-n-dn* plane. An upward dip of 10°, which would account for the dip of the mirror plane in Figure 11, would not be unexpected for the velocity vector; the dip in Figure 12 is somewhat greater, as could be expected near the terminus. This interpretation evidently requires

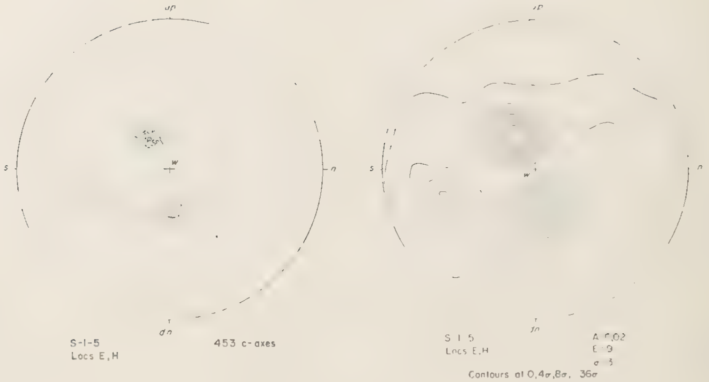


Fig. 13—Combination of the five fabric diagrams from Figure 11.

that the basic fabric pattern originate at some depth within the glacier, because at shallow depths, the pole of the free surface must be a principal stress axis, and this pole lies between *up* and *n* in the *up-n* plane, about 6° from the *up* axis. This suggestion is strengthened by the disposition of the two maxima in the (approximate) mirror plane: if the basic pattern as originally formed had mirror symmetry about a second mirror plane perpendicular to the first and passing through *ew*, as may be expected theoretically (see later discussion), the present position of the maxima indicates a rotation in the sense required by the shearing motion that has continued since the basic pattern developed.

Whether or not the above inferences prove to be correct, the fabrics as observed agree rather well with *Rigsby's* ideal diamond-shaped pattern [1951, p. 596]. The angles from the center of the pattern (as distinct from the pole of the projection) to the separate maxima are, on the average, about 45°, 40°, 23°, and 23° (to be compared with Rigsby's 45°, 40°, 26°, 26°). The long axis of the 'diamond' lies approximately in the direction of the stress vector acting across the persistent plane of maximum shear stress. The different degree of development of the near and far maxima is a prominent feature that does not, however, seem to be displayed in Rigsby's data. The much greater densities in the near maxima are displayed strikingly in the combined diagram S-1-5 in Figure 13. The far maximum between *w* and *n* (Fig. 11) is consistently weak, and in fact is not even resolved in the composite diagram; this lack of resolution ca be ascribed, however, to the unexplained loc variations in the orientation of the diamon shaped pattern. The weakness of this maximu may be responsible for Rigsby's recent concl sion (personal communication) that the mul ple-maximum fabrics are more typically cor posed of three rather than four maxima. T. fourth maximum persists, however, in the Bl Glacier data, though it appears to be easily lo perhaps through inadequate sampling.

Comparison of Theory, Observation, and Experiment

The evidence discussed above suggests tl the fabrics of coarse bubbly ice—in particu the multiple-maximum fabrics—originate in process of recrystallization from finer-textu ice. The fabric data from Blue Glacier ma possible an evaluation of recent theoretical a experimental studies of the recrystallizat process. The pertinent results of these stuc are as follows.

Recrystallization fabric theory of Ka [1959a, pp. 166–169]—Using as a basis Gib thermodynamic foundation, it is shown that der a stress situation corresponding to a hyd static pressure p superimposed upon a sh stress τ acting in direction A across the pl Ay having pole C (Fig. 14), the developmen preferred orientation in a recrystallizing ag gate of ice crystals is controlled, if the pro is not governed by differences in the inte condition of the crystals due to plastic st

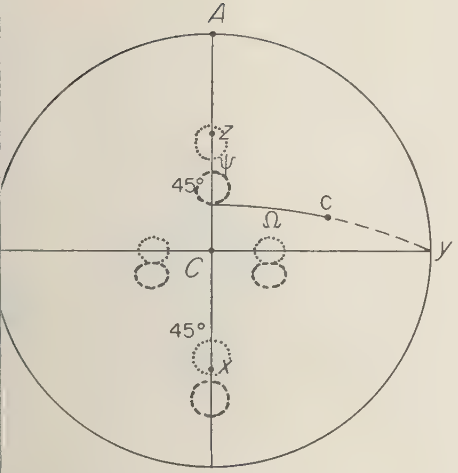


FIG. 14—Stress situation in glacier ice. The dia-ram shows a hemisphere, centered about the pole ' of the macroscopic foliation plane, in Schmidt azimuthal equal area) projection. x, y, and z are ie orientations of the principal stresses axes. The rincipal stresses are

$$\tau_{xx} = -\bar{p} + \tau,\ \tau_{yy} = -\bar{p},\ \tau_{zz} = \bar{p} - \tau.$$

. is the direction of maximum shear stress across ie foliation plane Ay. c represents an arbitrary ·e crystal c-axis orientation, defined by angles ψ nd Ω, the arc cy being a great circle through c and . Approximate location of observed maxima in -axis orientation density is shown by the dashed ircles, and the inferred location at time of fabric rigin by the dotted circles.

'ork, by the following combinations of the sin-le-crystal compliance constants:

$$= \tfrac{1}{2}(s_{11} + s_{33} - 2s_{13} - s_{44})$$
$$= -(5.3 \pm 0.4) \times 10^{-13}\ \mathrm{cm}^2\ \mathrm{dyne}^{-1}$$
$$= s_{11} - s_{12} - \tfrac{1}{2}s_{44}$$
$$= -(2.0 \pm 0.3) \times 10^{-13}$$
$$= -s_{11} + s_{33} - s_{12} + s_{13}$$
$$= (0.5 \pm 0.3) \times 10^{-13}$$

'he values and estimated errors are obtained rom the data of *Bass and others,* [1957], extra-olated to 0°C. The uncertainty in the values : probably somewhat larger than stated, to udge by the degree of disagreement between the data of *Bass and others* [1957] and of *Jona and Scherrer* [1952]. Since β and γ are negative, the preferred orientation for ice c-axes is at x, y, and z in Figure 14. Since ϵ is small compared with $|\beta|$ and $|\gamma|$, the expected preferred orientation does not depend upon whether or not diffusion or other transport of material along grain boundaries is operative. Since $\beta/\gamma > 1$, orientations at x and z are more stable than orientations at y. If it is assumed that the recrystallized fabric develops by the growth of more favorably oriented crystals at the expense of less favorably oriented ones, beginning with an initial fabric like that of fine ice, the predicted recrystallization fabric consists of maxima at x and z, connected by a weaker girdle extending between C and y and between C and $-y$.

In the case of uniaxial compression or tension, an evaluation of the appropriate combinations of elastic constants in equations (24), (25), (26), (27), (38), (39), and (40) [*Kamb,* 1959a] shows that ice classifies as 'type III,' regardless of whether or not diffusion transfer operates in the recrystallization mechanism. Thus, the c-axes of recrystallized ice should tend to cluster about the unique stress axis, without regard to whether it is compressional or tensional.

Theory of MacDonald [1959]—In this theory a Gibbs free energy for each crystal is identified with minus the elastic strain energy of the crysctal. *Brace* (to be published) has applied the theory to ice for certain special cases of applied stress, but I shall give a more general treatment here. The theories of Kamb and of MacDonald can be easily compared by writing the strain energy (per unit volume) U in the form (with summation convention)

$$U = \tfrac{1}{2}\tau_{ij}e_{ij} = \tfrac{1}{2}\tau_{ij}'\tau_{kl}'s_{ijkl}' - \bar{p}\tau_{ij}'s_{ijkk}' + U_0 \qquad (1)$$

where the e_{ij}, τ_{ij}, and τ_{ij}' are the components, respectively, of the strain, stress, and stress-deviator tensors, where $\bar{p}$ is the mean pressure, where the s_{ijkl}' are the four-index crystal compliance constants referred to the particular (primed) set of cartesian axes chosen, and where U_0 is a constant independent of crystal orientation (see *Brace,* to be published). The first term on the right side of (1) is simply $(\bar{u}-u^e)/V_0$ [*Kamb,* 1959a, equation (36)]. Thus if g is the Gibbs free energy per unit volume according to MacDonald,

$$g = -\bar{\mu}/V_0 + \bar{p}\tau_{ij}'s_{ijkl}' + g_0 \qquad (2)$$

where g_0 is a constant independent of orientation. The second term can be easily evaluated if principal axes are chosen. For a hexagonal crystal (c-axis called axis 3) we find that, for any particular value of α (*no sum over* α),

$$s_{\alpha\alpha kk}' = s_{\alpha\alpha kk}$$
$$+ (s_{33} - s_{11} + s_{13} - s_{12})l_{\alpha 3}^2 \qquad (3)$$

where $l_{\alpha\beta}$ is the direction cosine of the αth primed axis with respect to the c-axis of the crystal. Thus from (2) and (3) we find that for the stress conditions of Figure 14, and for a c-axis orientation defined by ψ, Ω (Fig. 14),

$$g(\psi, \Omega) - g(\pi/4, 0)$$
$$= -(1/V_0)[\bar{\mu}(\psi, \Omega) - \bar{\mu}(\pi/4, 0)]$$
$$- \bar{p}\tau(s_{33} - s_{11} + s_{13} - s_{12}) \cos 2\psi \cos^2 \Omega \quad (4)$$

where $\bar{\mu}(\psi, \Omega) - \bar{\mu}(\pi/4, 0)$ is given by equation (44) [*Kamb,* 1959a]. Similarly, under a uniaxial compression $\sigma = p_z - p_x$, where z is the unique stress axis and θ is the angle between z and the c-axis of the hexagonal crystal, we find

$$g(\theta) - g(0) = -(1/V_0)[\bar{\mu}(\theta) - \bar{\mu}(0)]$$
$$+ \bar{p}\sigma(s_{33} - s_{11} + s_{13} - s_{12}) \sin^2 \theta \qquad (5)$$

where $\bar{\mu}(\theta) - \bar{\mu}(0)$ is given by equation (37) [*Kamb,* 1959a].

From (4) and (5) it is seen that the fabric predictions of MacDonald's theory are just opposite to the predictions of Kamb's theory, except for the effect of the term containing the small coefficient $\epsilon = s_{33} - s_{11} + s_{13} - s_{12}$. For a pressure $\bar{p}$ less than 10 times the magnitude of the shear stress τ or the uniaxial compression σ, the correction term does not dominate. This verifies the particular results obtained by *Brace* (to be published) and shows that at reasonably low pressures the theory predicts preferred orientations at C and A in Figure 14, and under uniaxial compression or tension it predicts a small-circle girdle of c-axes inclined about 55° to the unique stress axis.

Experimental studies—In extensive experiments, *Steinemann* [1958a, b] has produced ice

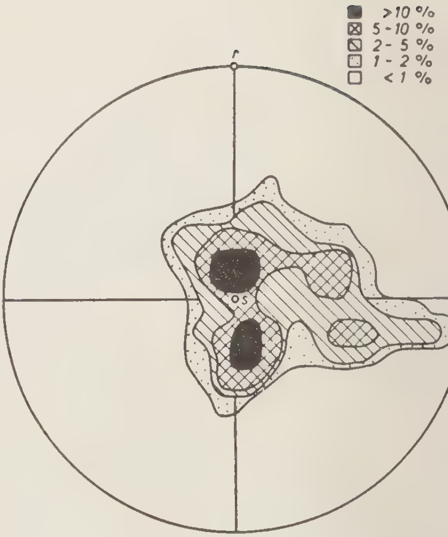


FIG. 15—Stereographic projection of the fabr of a torsion-shear experiment immediately aft unloading [from *Steinemann,* 1958b, Fig. 6 S is the pole of the shear plane, corresponding C in Figure 14, and r corresponds to y in Figure

recrystallization under stress in the laborator and *Shoumsky* [1958] has also produced r crystallization experimentally.

In Shoumsky's experiments recrystallizati under stress caused a reduction in grain size a led to a fabric and texture resembling that fine ice.

Steinemann's experiments produced a grow in crystal size during recrystallization und load ('primary parakinematic recrystallization For experiments under simple shear (torsion a hollow cylinder), only one fabric diagram, r produced here in Figure 15, is reported [*Stein mann,* 1958b, Fig. 60]. It is similar to t diagrams for natural fabrics, particularly in t general location of the two near maxima, though these maxima are not so well resolv from one another as in the natural fabrics a they are somewhat too close (13° and 22°) C (Fig. 14). A third maximum, in the AC pla about 50° from C, is weakly and irregularly veloped. In the experiment the sense of sh stress across the Ay plane was not determi in relation to the measured fabric, so that it not possible to identify this third peak for c in with a particular one of the far maxima bserved in glacier ice. However, the strength f the third peak in relation to the other two d its distance from C suggest that it corretes with the maximum between w and s in igure 13. A fourth maximum, corresponding to e weakest maximum observed in glacier ice, entirely lacking in the experimental fabric. *teinemann* [1958b, p. 49] apparently attributes e fourth maximum observed by *Rigsby* [1951] ther to observational errors (insufficient conol on the orientations of the thin sections) or episodes of shear of alternately opposite nse in the strain history of the material. The resent observations, however, discount both of ese interpretations and indicate that the urth maximum is a weak but real feature of e naturally occurring patterns.

Recrystallization after unloading ('postkineatic recrystallization') of a specimen stressed simple shear led to a single maximum in axis density centered about the pole of the ear plane [*Steinemann,* 1958b, Fig. 61].

Recrystallization under uniaxial compression d to rather weak fabrics that are difficult to sess because they do not exhibit rotational mmetry about the compression axis [*Steineann,* 1958b, Figs. 62–64]. If this difficulty is sregarded and the experimental fabrics are roted about the compression axis so as to achieve tational symmetry artificially, the orientations tline a diffuse small-circle girdle of about 30° dius. The difference between a diffuse smallrcle girdle and a simple broad maximum cenred about the compression axis is small for ost of the reported data and may or may not statistically significant.

Postkinematic recrystallization after loading uniaxial compression led to fabrics which do t differ greatly from those developed under ad.

Discussion—The limited experimental data r recrystallization fabrics developed under nple shear are in fair enough agreement with e observed data from Blue Glacier that it ems likely that the same process is involved the experimental and natural recrystallization. e interdigitating texture developed experientally at not too high strain rates [*Steineann,* 1958b, pp. 40–41] adds a second strong nilarity between the natural and experimental processes. Such differences of fabric as exist may be due to the relatively small amount of total strain (a few per cent) involved in the experiments, and the (perhaps related) relatively small size of the recrystallized grains (largest grains 0.5 to 1 cm in diameter).

Likewise, the close similarity between fabric and texture of fine-ice layers in Blue Glacier and the 'shear zones' produced in *Shoumsky's* [1958] experiments leaves little doubt that the experimental and natural processes here are the same.

For comparison with recrystallization fabrics developed experimentally under uniaxial compression it is not possible at present to identify ice within Blue Glacier that demonstrably has had a strain history of uniaxial compression. However, ice at locality G may have had such a history, and the fabric found there (C-8, not illustrated here) for coarse bubbly ice partially (90%) recrystallized from fine ice is a single broad maximum in c-axis density centered about the axis of (presumed) compression. These data are discussed in more detail separately [*Allen and others,* to be published].

Theoretical interpretation of ice fabrics is in a less satisfactory stage of development than is experimental reproduction of the natural patterns. It would seem reasonable to expect that the recrystallization process that produces the complex texture of coarse bubbly ice should give rise to a different type of preferred orientation than is achieved in the simple plastic-flow process that accounts for the fabrics of sheared polar ice and of fine temperate ice. In the latter, the role of recrystallization is perhaps mainly to allow polygonalization of crystals that have become bent in the plastic flow process. If this is so, then the theories of recrystallization-fabric origin summarized above are not applicable to fine ice.

The position of the maxima in the coarse-ice four-maximum fabrics is plotted in Figure 14 (in interpreting the observed fabrics it is assumed, as discussed above, that the monoclinic symmetry of the fabric must fit the stress symmetry). If the post-recrystallization rotation (envisaged above) were removed, the maxima would lie approximately in the dotted positions as shown. The far maxima would then be accounted for approximately by Kamb's theory, but there is no explanation for the stronger near maxima. MacDonald's theory predicts maxima at C and A, where they are not observed. Both theories assume homogeneous stress, and it appears that a removal of this doubtless unrealistic assumption may bring their predictions closer together; it would not, however, remove the basic incompatibility of the two approaches. One of the two maxima predicted by MacDonald's theory, on the other hand, fits the fine-ice fabrics and the corresponding experimental recrystallization fabrics if they are assumed to be produced by the mechanism described by the theory (which, however, seems doubtful). A small-circle girdle of the kind predicted for uniaxial compression or tension by MacDonald's theory has not been observed in glacier ice, but the experimental results for this case (to the extent that they are acceptable) indicate a small-circle girdle, but of radius rather smaller than predicted by the theory (30° vs. 55°).

The degree of disagreement between the theoretically predicted fabrics and the observed fabrics is sufficient to make both theories unacceptable. I think that the shortcomings of my theoretical treatment [*Kamb*, 1959a] lie not in the underlying thermodynamic principles but in the simplicity of the model assumed. What is needed is a further refinement that takes into account (1) the non-homogeneous stress distribution, (2) the thermodynamic effect of plastic strain work, and (3) the direct reorientation accompanying intracrystalline plastic flow.

Acknowledgments—The petrofabric studies reported here are a part of the Blue Glacier research program of the California Institute of Technology, carried out in conjunction with the International Geophysical Year, and supported by the U. S. National Committee for the IGY. The field data were obtained during August 1957 and August 1958. I am indebted to R. P. Sharp, director of the Blue Glacier project, for much helpful advice and discussion and for the opportunity to carry out the work, and to the other members of the research party for help and support in innumerable ways.

Appendix: Preparation of Orientation-Density Diagrams

The primary question in interpreting any fabric diagram is whether or not the diagram shows statistically significant preferred orientation. Statistical methods of testing significance of preferred orientation have been discussed by *Chayes* [1949] and *Flinn* [1958]. However, instead of using numerical methods, I have taken a simpler and more graphic approach that utilizes the orientation-density diagram to exhibi directly the statistical significance of the data The density diagram is treated here not simply as a means of portraying the original orientation data, as is done in conventional contouring procedures [*Fairbairn*, 1949, pp. 285–290]. Instead the actual data are presented as scatter diagrams, which contain the information that would be needed for a statistical test by numerica methods, and the density diagrams are prepared in such a way as to abstract from the scatte diagrams the orientation information that is o statistical significance.

The measure of statistical significance is th probability that the observed orientation density could have resulted from random samplin of a population that lacks preferred orientation To control this probability, the area A of th counter used in the conventional (Schmidt contouring procedure is so chosen that, if th population lacks preferred orientation, the number of points E expected to fall within a give area A is three times the standard deviation the number of points n that will actually fa within the area under random sampling of th population. This insures that the observed orientation densities, if obtained by random samplin of a non-preferentially oriented population, w not fluctuate wildly from the expected densi E/A. Observed densities that differ from $E/$ by more than two or three times the standa deviation σ (for random orientation) are th likely to be significant, and the more so if t significantly higher densities are clustered in o portion of the diagram. The observed densit are therefore contoured in intervals of 2σ, the values 0, 2σ, 4σ, etc., the expected densi E for no preferred orientation being 3σ.

For a given area A, expressed as its fracti of the area of the hemisphere, the distributi of n values for random samples of size N binomial, and for a population without p ferred orientation we find

$$\sigma/E = \sqrt{(1 - A)/NA}$$

where $E = NA$.

etting $\sigma/E = 1/3$, we compute for a given bric with N points the appropriate area A of e counter to be used in preparing the density agram. The pertinent values accompany each nsity diagram presented.

Diagrams prepared in this way have a conderably smoothed appearance in comparison ith conventional density diagrams. This is beuse most of the irregular detail visible in conntional diagrams is of no statistical signifince, the conventional choice $A = 0.01$ being ually much too small. With the choice of A ed here, the difference between the Schmidt d Mellis contouring methods [*Flinn,* 1958] n lead to no statistically significant differences the positions of the contours and may therere be disregarded; the Schmidt method is, of urse, much easier to use.

A precise statistical evaluation of density diaams prepared as described above can be cared further, and the levels of significance formuted, but such refinements are not necessary re.

References

LLEN, C. R., W. B. KAMB, M. F. MEIER, AND R. P. SHARP, Structural features of Blue Glacier, Washington (to be published).

ADER, H., Introduction to ice petrofabrics, *J. Geol., 59,* 519–536, 1951.

ASS, R., D. ROSSBERG, AND G. ZIEGLER, Die elastischen Konstanten des Eises, *Z. Physik, 149,* 199–203, 1957.

RACE, W. F., Orientation of anisotropic minerals in a stress field: discussion, (to be published).

HAYES, F., Statisticals analysis of three dimensional fabric diagrams, *in* Fairbairn, H. W., *Structural Petrology of Deformed Rocks,* Addison-Wesley Press, Cambridge, Mass., 308–326, 1949.

STEIN, S., AND R. P. SHARP, Oxygen isotope studies, *IGY Bull., 21,* 9–12, 1959; also *Trans. Am. Geophys. Union, 40,* 81-84, 1959.

IRBAIRN, H. W., *Structural Petrology of Deformed Rocks,* Addison-Wesley Press, Cambridge, Mass., 344 pp., 1949.

INN, D., On tests of significance of preferred orientation in three-dimensional fabric diagrams, *J. Geol., 66,* 526, 1958.

NA, F., AND P. SCHERRER, *Helv. Phys. Acta, 25,* 35, 1952.

KAMB, W. B., Theory of preferred crystal orientation developed by crystallization under stress, *J. Geol., 67,* 153–170, 1959a.

KAMB, W. B., The correction of universal stage orientation (abstr.), A. C. A. Summer Meeting, Ithaca, N. Y., 1959b.

LANGWAY, C. E., Ice fabrics and the universal stage, *SIRPRE Tech. Rept. 62.,* 1958.

MACDONALD, G. J. F., The orientation of anisotropic minerals in a stress field, *Geol. Soc. Am. Spec. Paper* (in press), 1959.

MEIER, M. F., *Mode of flow of Saskatchewan Glacier, Alberta, Canada,* Thesis, Calif. Inst. of Technology, 1957.

MEIER, M. F., G. P. RIGSBY, AND R. P. SHARP, Preliminary data from Saskatchewan Glacier, Alberta, Canada, *Arctic, 7,* 3–26, 1954.

NAKAYA, U., The deformation of single crystals of ice, *Symp. de Chamonix, Assoç. Intern. d'Hydrol. Sci., Publ. 47,* 229–240, 1958.

PERUTZ, M. F., AND G. SELIGMAN, A crystallographic investigation of glacier structure and the mechanism of glacier flow, *Proc. Roy. Soc. A, 172,* 335–360, 1939.

RIGSBY, G. P., Crystal fabric studies on Emmons Glacier, Mt. Ranier, Wash., *J. Geol., 59,* 590–598, 1951.

RIGSBY, G. P., *Studies of crystal fabrics and structures in glaciers,* Thesis, Calif. Inst. of Technology, 1953.

RIGSBY, G. P., Study of ice fabrics, Thule area, Greenland. *SIPRE Report 26,* 1955.

RIGSBY, G. P., Fabrics of glacier and laboratory deformed ice, *Symp. de Chamonix, Assoc. Intern. d'Hydrol. Sci., Publ. 47,* 351–358, 1958.

SCHWARZACHER, W., AND N. UNTERSTEINER, Zum Problem der Bänderung des Gletschereises, *Sitzber. Öster. Akad. Wiss., Math.-Naturw. Klasse, Abt. IIa, 162,* 111–145, 1953.

SHOUMSKY, P. A., The mechanism of ice straining and its recrystallization, *Symp. de Chamonix, Assoc. Intern. d'Hydrol. Sci., Publ. 47,* 244–248, 1958.

STEINEMANN, S., Results of preliminary experiments on the plasticity of ice crystals, *J. Glaciol., 2,* 404–412, 1954.

STEINEMANN, S., Resultats experimentaux sur la dynamique de la glace et leur correlations avec le mouvement et la petrographie des glaciers, *Symp. de Chamonix, Assoc. Intern. d'Hydrol. Sci., Publ. 47,* 184–198, 1958a.

STEINEMANN, S., Experimentelle Untersuchungen zur Plastizitat von Eis, *Beitr. Geol. Schweiz, Hydrolgie, no. 10,* esp. pp. 46-50, 1958b.

VON KLEBELSBERG, R., *Handbuch d. Gletscherkunde u. Glacialgeologie,* Springer Verlag, Vienna, 1948.

(Manuscript received June 26, 1959.)

Salt Intrusion into Fresh-Water Aquifers

Harold R. Henry

U. S. Geological Survey and Michigan State University
East Lansing, Michigan

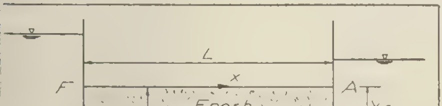
Abstract—In a coastal aquifer a steady flow of fresh water toward the sea can limit the encroachment of the salt water into the aquifer. This action is treated on the assumptions that the flow is steady and two-dimensional, that the salt and fresh water are immiscible, and that there is no fingering. Theoretical equations for the shape and location of the interface and for the boundary velocities are derived for several sets of boundary conditions.

The uncertainty of the location of the interface is circumvented by use of a hodograph plane. In addition, a complex potential is employed and related to the hodograph by conformal mapping.

Certain boundary conditions represent inversions of gravity seepage through dams for which solutions already exist. Numerical computations are also presented for a semi-infinite aquifer having a vertical seepage face and one having a horizontal seepage face.

Introduction—The location of the interface
tween the salt and fresh water in a coastal
uifer is of obvious interest in the practical
e of ground water. The seaward flow of fresh
ter in the aquifer is necessary to keep salt
ter from contaminating the aquifer. As a first
proximation to this action, exact solutions
ve been obtained for several boundary con-
ions, on the assumption that the salt and
sh water are immiscible and that there is no
gering and no tidal action. The effects of dif-
sion and tidal action result in a 'zone of dif-
sion' which will be investigated as a continua-
n of the present study.

The configurations analyzed are shown in
gures 1 and 6. In Figure 1 the seepage face
} is vertical, and in Figure 6 it is horizontal.
each case the problem is solved first for a
ite aquifer as shown. The equations are subsequently modified to solve the cases where the aquifer extends to infinity to the left. Another limiting case which is treated is that of an aquifer that is infinite in extent to the left and infinite in depth also.



ı. 1—Case I, horizontal confined aquifer with vertical outflow face.

Notation—The following symbols are used.

A, B, C, D, E, F	points on the boundary of the aquifer
a, b, c, d, e	coordinates of points in the complex plane
$f(z)$	complex velocity potential $= \phi + i\psi$
$f'(z)$	first derivative of f with respect to z
$f''(z)$	second derivative of f with respect to z
h	piezometric head, in feet of fresh water
i	$\sqrt{-1}$
J	the elliptic modular function
K	complete elliptic integral of first kind
K'	complete elliptic integral of first kind, with complementary modulus
k	modulus of elliptic integral
$\bar{k}$	permeability of aquifer
k'	product of permeability and buoyancy $= \bar{k}(\gamma_s - \gamma_0)/\gamma_0$
L	length of aquifer
Q	discharge per unit width
q	complex velocity $= u + iv$
$\bar{q}$	conjugate complex velocity $= u - iv$
q_1, q_2	transformations of the hodograph
S	depth of aquifer
λ, w, t	complex variables

u	velocity in x direction
v	velocity in y direction
x	horizontal coordinate
y	vertical coordinate measured downward
z	complex variable
α	the integral, $-\dfrac{3}{2\pi}\displaystyle\int_0^1 \dfrac{\beta(t)}{t-\lambda}\,dt$
β	central angle in hodograph plane
γ_s	specific weight of salt water
γ_0	specific weight of fresh water
ϕ	velocity potential
ψ	stream function
$\tau + i\theta$	the auxiliary variable

Governing Equations and Boundary Conditions—The use of Darcy's law for a homogenous two-dimensional aquifer,

$$u = -\partial\phi/\partial x, \qquad v = -\partial\phi/\partial y \qquad (1)$$

in the continuity equation,

$$\partial u/\partial x + \partial v/\partial y = 0 \qquad (2)$$

yields Laplace's equation for ϕ:

$$\nabla^2\phi = 0 \qquad (3)$$

A stream function ψ orthogonal to ϕ can be defined by the Cauchy-Riemann equations and a complex velocity potential $f(z)$ written as

$$f(z) = \phi + i\psi \qquad (4)$$

The conjugate of the complex velocity is defined by

$$\bar{q} = u - iv = -f'(z) \qquad (5)$$

The variation of ϕ along the interface of salt and fresh water and along the seepage surface is controlled by the pressure in the salt water, which is assumed to be at rest. Thus along AB and BD in Figures 1 and 6

$$\phi = \phi_0 + k'y \qquad (6)$$

where ϕ_0 is constant and y is measured downward. Differentiation of ϕ with respect to the distance r along the surface of seepage AB yields

$$-\partial\phi/\partial r = u_r = u\cos\alpha + v\sin\alpha \qquad (7)$$

where α is the angle between the seepage surface and the horizontal; α is $\pi/2$ in Figure 1 and

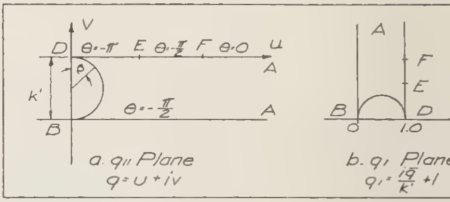


Fig. 2—Hodograph for case I.

zero in Figure 6. Substitution of (6) into (
yields

$$k'\sin\alpha + u\cos\alpha + v\sin\alpha = 0 \qquad ($$

This is the equation of the line in the hodograph plane (u,v) representing the velocity AB. (See Figs. 2a and 7a).

Differentiation of (6) with respect to distan s along the interface BD yields

$$\partial\phi/\partial s = -k'(\partial y/\partial s) \qquad ($$

Since the interface is a streamline, $\partial\phi/\partial s$ is t total velocity $(u^2 + v^2)^{1/2}$. Multiplying (9) l this factor gives

$$u^2 + v^2 + k'v = 0 \qquad (1$$

Equation (10) is a circle in the hodograph q plane.

The hodograph for a flow boundary that is straight line of constant potential is a straig line passing through the origin and perpendicu to the real boundary. For a straight-flow bour ary that is a streamline, the hodograph l passes through the origin parallel to the r boundary.

The problem is solved in the following ma ner. First, the velocities on the boundaries the flow system are mapped on the hodogra plane. In general u and v or a relationship tween u and v is required for every point the boundary. Next, the complex potential the boundaries of the flow system is determin Now if the hodograph diagram is transforr into the f diagram by some function, this fu tion can be substituted directly into (5), wh can subsequently be integrated to obtain solution. It is not always feasible to transf f directly into q, and it may be convenient employ auxiliary planes to effect the transfor tion in several steps. Further, if f is not kn

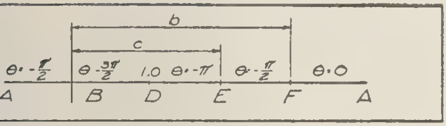


FIG. 3—The λ plane; $\lambda(q_1) = J(1 - 1/q_1)$.

ong some portion of the boundary, an auxiliary nction of f may be introduced to make the oblem tractable.

Case I. Vertical outflow face—The area BDEF (Fig. 1) is saturated with fresh water. he area *BCD* contains salt water at rest. The companying hodograph is shown in Figure 2a. he q_1 plane (Fig. 2b) is a modified hodograph efined by

$$q_1 = (i/k')\bar{q} + 1 \qquad (11)$$

The zero-angle triangle, with the semicircular de, in the q_1 plane (Fig. 2b), can be transrmed into the upper half-plane shown in Fige 3 by use of the elliptic modular function (q_2) as defined by *Nehari* [1952] and by *hnke and Emde* [1945]. Using

$$q_2 = 1 - 1/q_1 \qquad (12)$$

e have

$$\lambda(q_1) = J(q_2) \qquad (13a)$$

ne values of λ corresponding to the flow oundaries can be computed by using the inrse of the elliptic modular function:

$$1 - \frac{1}{q_1} = \frac{iK'(\lambda)}{K(\lambda)} \equiv i\,\frac{K'}{K}\,(\lambda) \qquad (13b)$$

lues of K'/K for λ real and between zero and ity are given by *Hayashi* [1930]. Thus, within is range, arbitrary values of λ may be subtuted into (13b) and corresponding values of computed. For other ranges of λ simple linear ansformations are required to permit use of e tabular values.

r $-\infty < \lambda \leq 0$ one may use

$$1/(1 - \lambda) = J(q_1) \qquad (14a)$$

$$q_1 = i\,\frac{K'}{K}\left(\frac{1}{1 - \lambda}\right) \qquad (14b)$$

$1 \leq \lambda < \infty$ one may use

$$1/\lambda = J(q_1 - 1) \qquad (15a)$$

$$q_1 - 1 = \frac{iK'}{K}\left(\frac{1}{\lambda}\right) \qquad (15b)$$

Equations 13a or b, 14a or b, and 15a or b may be considered alternate forms of the same transformation between q_1 and λ.

To complete the solution of the problem, it is necessary to express the complex potential f, or some function of f, in terms of $\bar{q}$ so that equation (5) may be integrated. Because the distribution of ψ on *AB* in Figure 1 is unknown, the method of *Hamel* [1934] is followed. Use is made of the function

$$\tau + i\theta = -\ln\,[-f''(z)] \qquad (16)$$

in which

$$-f''(z) = \frac{du - idv}{dx + idy} \qquad (17)$$

and $(-\theta)$ is the argument of $-f''(z)$. From (17) and the hodograph it follows that

$$\begin{aligned} \theta_{AB} &= -\pi/2; \quad \theta_{BD} = -3\beta/2 \\ \theta_{DE} &= -\pi; \quad \theta_{EF} = -\pi/2; \quad \theta_{FA} = 0 \end{aligned} \qquad (18)$$

The angle β in the second of equations (18) is indicated in the hodograph in Figure 2.

Referring to Figure 3, we see that the values of θ are known at all points of the real axis of the λ plane. Values of θ may then be obtained for all points of the upper half plane by a Fourier integral solution. Subsequently values of τ are obtained by integrating the Cauchy-Riemann equations for τ and θ. *Muskat* [1937] gives details of this procedure which leads to the generalized Poisson formula:

$$\tau + i\theta = \tau_0 + \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\theta(t)(\lambda t + 1)\,dt}{(t - \lambda)(1 + t^2)} \qquad (19)$$

where τ_0 is an arbitrary constant.

Evaluating the integral in (19) and applying the results in (16) yields, after solving for $-f''$,

$$-f'' = \frac{-\sqrt{\lambda(c - \lambda)(b - \lambda)}}{1 - \lambda} \cdot \exp\left(-\tau_0 + \frac{3}{2\pi}\int_0^1 \frac{\beta(t)}{t - \lambda}\,dt\right) \qquad (20)$$

where b and c are the coordinates of points E and F respectively in the λ plane. Differentia-

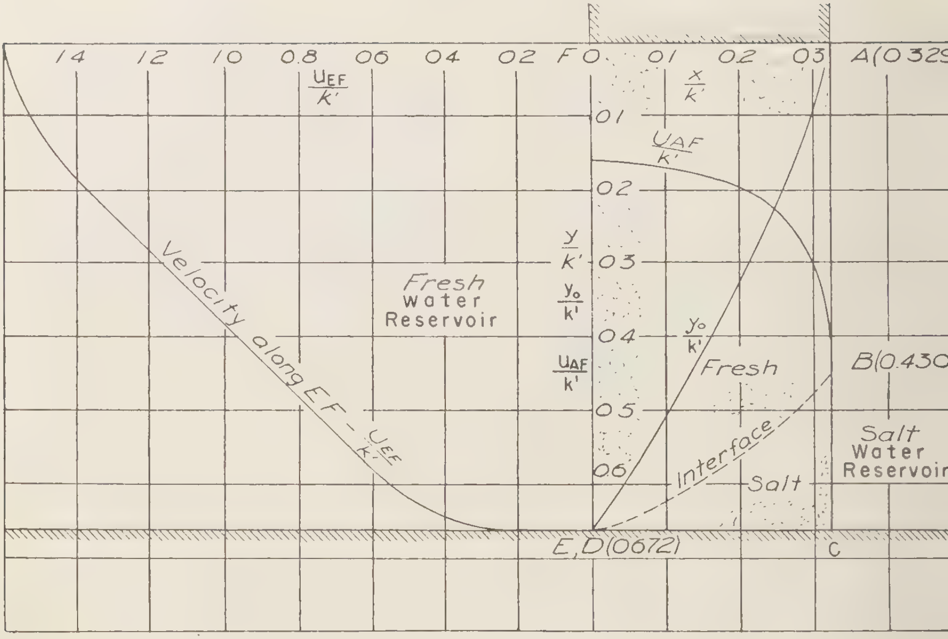


FIG. 4—Salt intrusion under a short sand formation (An inversion of Muskat's case III; curves a identical to Muskat's).

tion of (5) and substitution into (20) gives, after solving for z,

$$z = c_1 \int e^{\tau + i\theta}\, d\bar{q} + C_2 \qquad (21)$$

where

$$e^{\tau} = \frac{1 - \lambda}{\sqrt{\lambda(c - \lambda)(b - \lambda)}}\, e^{\alpha} \qquad (22)$$

and

$$e^{\alpha} = \exp\left(-\frac{3}{2\pi}\int_0^1 \frac{\beta(t)}{t - \lambda}\, dt\right) \qquad (23)$$

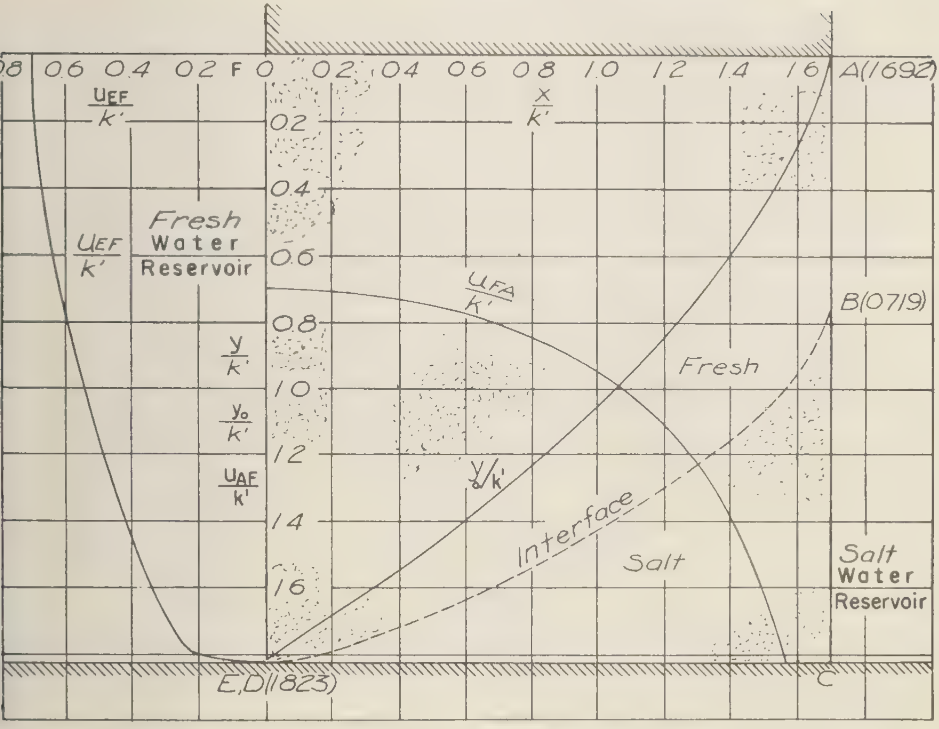
Hamel and Gunther [1935] and *Muskat* [1935] give tabulated and graphical values of e^{α} as a function of λ. Thus, the velocities at the boundaries of the flow and the location of the interface may be computed by substituting arbitrary real values of λ into the proper equation, (13), (14), or (15), and by subsequent graphical or numerical integration of (21).

If the relative positions of the levels of the fresh water and the salt water in Figure 1 a such that point D coincides with E—that is, t interface reaches to the upstream reservoir then $\phi_E = \phi_D$ and

$$e^{\tau} = \sqrt{\frac{1 - \lambda}{\lambda(b - \lambda)}}\, e^{\alpha} \qquad (24$$

This case is mathematically identical with t case of seepage through a rectangular dam wi no tailwater in which the permeability k h been replaced by k'. *Muskat* [1935] gives n merical solutions for the discharge Q and t height of seepage surfaces AB for four su cases. He plots computed boundary velociti for two cases and also draws in the free surfa without calculation, following the general fe tures of a previous case with tailwater compute by *Hamel and Gunther* [1935]. Muskat's curv have been inverted and relabeled to fit the pre ent application, as shown in Figures 4 and 5.

When the aquifer extends an infinite distan to the left in Figure 1 the point E in the hod graph coincides with point F and e^{τ} becomes

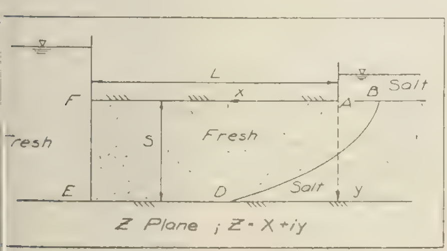


IG. 5—Salt intrusion under a short sand formation (An inversion of Muskat's case VI; curves are identical to Muskat's).

$$e^{\tau} = \frac{1-\lambda}{(c-\lambda)\sqrt{\lambda}}\,e^{\alpha} \qquad (25)$$

oreover, if the aquifer also extends an infinite stance downward, the hodograph points D, E, d F coincide and e^{τ} becomes

$$e^{\tau} = (1/\sqrt{\lambda})e^{\alpha} \qquad (26)$$



G. 6—Case II, horizontal confined aquifer with horizontal outflow face.

The numerical solution for this case is given in the last section of this paper.

Case II. Horizontal outflow face—Figure 6 represents a flow in which the seepage face AB is horizontal. (*Polubarinova-Kochina* [1940] discusses the unsteady characteristics of a similar case.) The hodograph or q plane for the present case is shown in Figure 7a and a transformed hodograph (q_1 plane) is shown in 7b. By use of the Schwarz-Christoffel theorem, the semi-infinite strip of the q_1 plane may be transformed to the upper half of the t plane of Figure 8 by

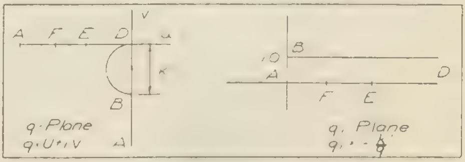


FIG. 7—Hodograph for case II.

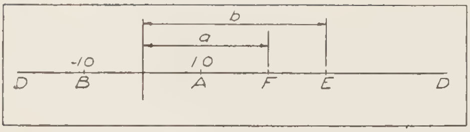


FIG. 8—The t plane; $t = \cosh \pi q_1 = \cosh - \pi k'/q$.

$$t = \cosh \pi q_1 = \cosh(-\pi k'/\bar{q}) \qquad (27a)$$

or inversely

$$\bar{q} = (-\pi k')/(\cosh^{-1} t) \qquad (27b)$$

In the complex potential plane f the boundaries of the flow form the rectangle shown in Figure 9, where the potential along the seepage face AB is taken as zero. The method of Hamel which introduces the function $\tau + i\,\theta$ defined by (16) could be employed in this case also with a result similar to equation (20). However, it would be tedious to compute the exponential factor similar to e^{α}. This difficulty is obviated by using a more direct method of mapping the rectangle in the f plane into the infinite strip of the q_1 plane by transforming each into the identical half-plane. This is accomplished by first mapping f into the upper half w plane of Figure 10, using the Schwarz-Christoffel theorem:

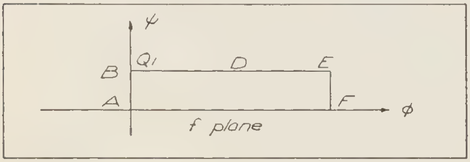


FIG. 9—Complex potential for case II.

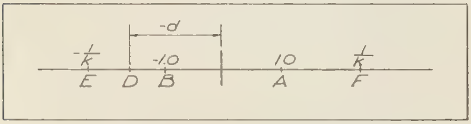


FIG. 10—w plane; $w = sn\ [(\text{if } 2/Q + 1)K]$.

$$w = sn[if(2/Q) + 1]K \qquad (28a)$$

or inversely

$$f = \frac{Q}{2i}\left(\frac{sn^{-1}w}{K} - 1\right) \qquad (28b)$$

where

$$sn^{-1}w = \int_0^w \frac{dw}{\sqrt{(1 - w^2)(1 - k^2w^2)}} \qquad \text{(28c}$$

and

$$K = sn^{-1}(1, k)$$

The t plane of Figure 3 can be mapped int the w plane by the linear transformation

$$w = \frac{td - 1}{d - t} \qquad \text{(29}$$

where $-d$ is the coordinate of point D in the plane. The relation between the coordinates and b of the t plane and the values d and k the w plane are found by substitution into (29

$$a = \frac{d + k}{kd + 1} \qquad \text{(30}$$

$$b = \frac{k - d}{kd - 1} \qquad \text{(3}$$

Thus the problem is solved in principle, since is expressed parametrically in terms of f throug (27b), (28b), and (29), and in principle (can be integrated to give

$$z = \int \frac{-df}{\bar{q}} + \text{const.} \qquad \text{(3}$$

The numerical solution of a particular ca requires given values of either a and b or d ar k. Assuming either pair of values corresponds effect to assigning values to the two dimensio less quantities $Q/k'S$ and L/S, where S is t vertical thickness of the aquifer and L is i length.

For the case of a horizontal strip of aquif semi-infinite in length (Fig. 6), the points and F are at an infinite distance to the left, t two corresponding points in the hodograph a coincident, and in the t plane, $a = b$. Also Figure 9, E and F would be at infinity to t right. In the w plane this requires that $k =$ and that the points E and F be at infinity. Th equation (28b) transforming the w plane in the f plane degenerates into

$$f = \frac{Q}{\pi} \cosh^{-1} w \qquad \text{(3}$$

ıe relation between the w and t planes is still pressed by (29), and (30) and (31) degenere into $a = b = d$. Numerically, a particular se is determined by assigning a value to either or d which in principle corresponds to a cerin value of the ratio $Q/k'S$.

Finally, if in Figure 6, in addition to L beming infinite, the vertical thickness S becomes finite also, the points D, E, and F will coincide the hodograph. Then, upon division by Q, ıe f plane becomes identical to the q_1 plane, elding

$$f/Q = -k'/\bar{q} \qquad (34)$$

or this case, (32) can be integrated to yield

$$z = f^2/2k'Q \qquad (35)$$

This is mathematically identical with a seepze problem solved by *Kozeny* [1953] for a ee-surface flow in a semi-infinite aquifer drainıg downward through a horizontal-outflow face ith k' substituted for $\bar{k}$. This correspondence etween the Kozeny solution and the salt-inusion problem for the semi-infinite aquifer has een noted and utilized by *Glover* [1959].

Computations for the semi-infinite cases I and —In order to compare the numerical results ır horizontal and vertical seepage faces, diıensionless plots are presented for each case. ; should be noted that all solutions of the semiıfinite problem, case I, are geometrically simiır. This is true also in the semi-infinite space case II. For case I (18) and (26) are substiıted into (21), for which values of $\bar{q}$ are comıted from the appropriate equation, (13), 4), or (15). Then (21) is integrated numeriılly or graphically. Along AB, λ varies from ∞ to zero and the corresponding values of $\bar{q}$ 'e computed from (14b), which in this case is

$$\frac{u}{k'} = \frac{K'}{K}\left(\frac{1}{1-\lambda}\right) \qquad (36)$$

alues of K'/K may be found in *Hayashi's* bles [1930]. Values of e^α are taken from *Hamel* *ıd Gunther* [1935]. The integration indicated ' (21) was performed by the trapezoidal rule to a value of $u = 2.8$ corresponding to $\lambda =$ 400. For larger negative values of λ the anatic approximation introduced by *Hamel* and *unther* [1935] was used

$$\frac{d\tilde{q}}{k'} \cong \frac{1}{\pi}\frac{d\lambda}{\lambda}\,; \qquad e^{\alpha} \cong 1 \qquad (37)$$

Computed depth y_s of the seepage surface AB is

$$\frac{y_s k'}{Q} = 0.741$$

as shown in Figure 11.

For the interface BD, equation (10) yields

$$\bar{q} = \frac{-ik'}{2}\left(e^{i\beta} - 1\right) \qquad (38)$$

and use of equations (18) and (26) yields

$$\int e^{\tau + i\theta}\, d\bar{q} = \int \frac{k'}{2}\frac{1}{\sqrt{\lambda}} e^{\alpha} e^{-i\beta/2}\, d\beta \qquad (39)$$

Solving for the x and y coordinates from equation (21) gives

$$\frac{k'(x - x_B)}{Q} = \frac{1}{2}\int_{\pi}^{\beta} \cos\frac{\beta}{2}\cdot\frac{1}{\sqrt{\lambda}}\, e^{\alpha}\, d\beta \qquad (40a)$$

$$\frac{k'(y - y_B)}{Q} = \frac{1}{2}\int_{\pi}^{\beta} \sin\frac{\beta}{2}\cdot\frac{1}{\sqrt{\lambda}}\, e^{\alpha}\, d\beta \qquad (40b)$$

Values of β corresponding to given values of λ were computed with (13b), which in view of (11) and (38) may be rewritten as

$$\tan \beta/2 = \frac{K'}{K}(\lambda) \qquad (41)$$

The integrals of equations (40a) and (40b) were computed by the trapezoidal rule to give the coordinates of the interface, shown by the solid line marked BD in Figure 11.

The piezometric head on the boundary AF is indicated by the line with ordinate $k'y_0/Q$ in Figure 11. The depth y_0 (x) would be the ordinate of the interface if hydrostatic conditions prevailed along vertical sections. This computation requires first the computation of u versus x as shown by the solid curve above AF. The latter proceeds from a graphical integration of (21), by use of (15) for the relation between λ and q_1. The head was subsequently computed by

$$h_x - h_A = y_0\left(\frac{\gamma_s - \gamma_0}{\gamma_0}\right) = \int \frac{u}{\bar{k}}\, dx \qquad (42a)$$

$$\frac{k'y_0}{Q} = \int \frac{u}{k'}\, d\left(\frac{xk'}{Q}\right) = \int \frac{u}{k'}\, e^{\tau}\, \frac{du}{k'} \qquad (42b)$$

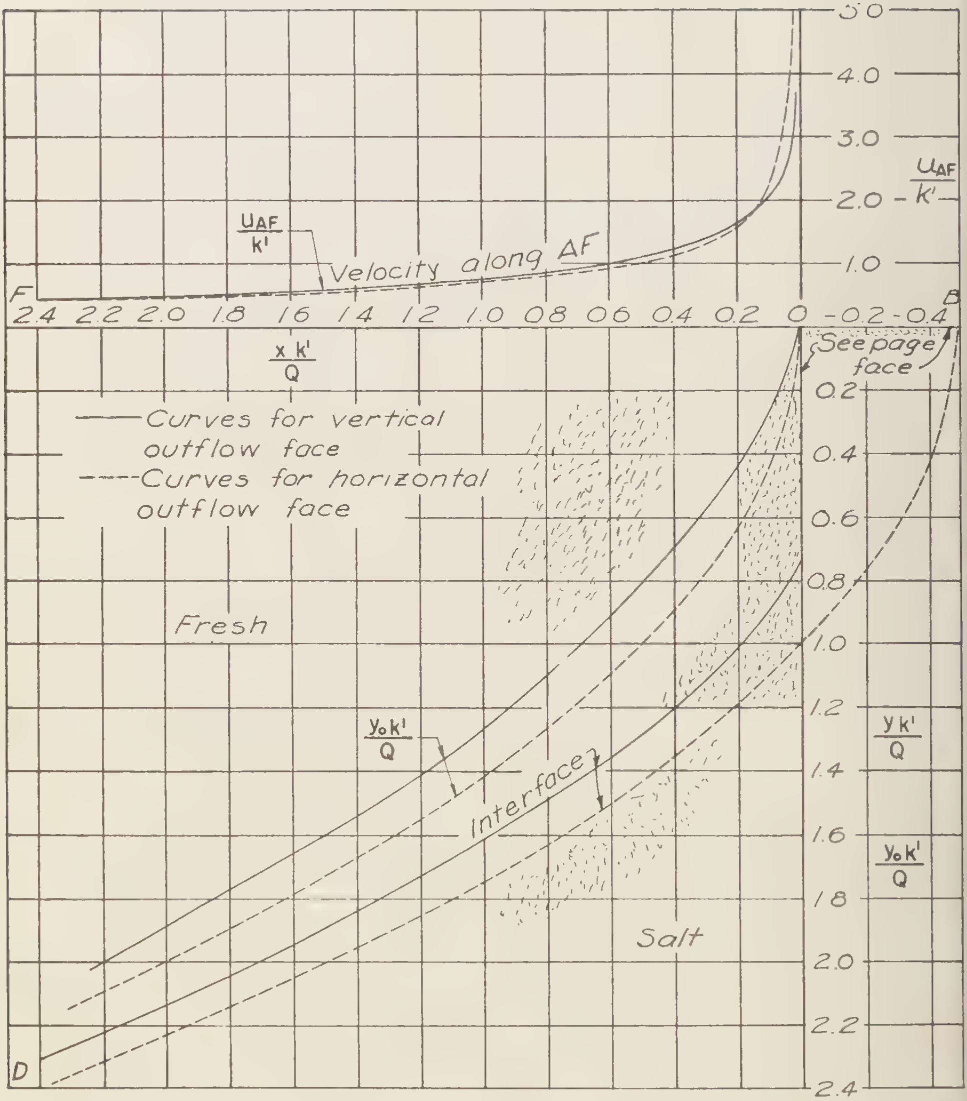


Fig. 11—Numerical solution of the semi-infinite aquifer for cases I and II.

in which h_x represents the head on AF at any abscissa and h_A represents the head at A in feet of fresh water. Equation (42b) was integrated graphically to produce the upper curve for $k'Y_0/Q$ shown in Figure 11.

Computations for case II with a horizontal-outflow face are much simpler. The width of the seepage surface is obtained by substituting $y = 0$, $x = x_s$, $\psi = Q$, and $\phi = 0$ into (35). The result is

$$x_s k'/Q = -\tfrac{1}{2} \qquad (43)$$

The interface is computed by substituting $\psi = Q$ and $\phi = k'y$ into (35).

$$\left(\frac{yk'}{Q}\right)^2 - 2\left(\frac{xk'}{Q}\right) - 1 = 0 \qquad (44)$$

The dashed curve BD in Figure 11 is a plot of (44).

The piezometric head on AF, indicated by y_0, is obtained by substituting $\psi = 0$ and $y = 0$ into (35) and solving for y_0.

$$\frac{y_0k'}{Q} = \sqrt{2\left(\frac{xk'}{Q}\right)} \qquad (45)$$

The velocity along AF is given by differentiating ϕ.

$$\frac{u}{k'} = -\frac{1}{k'}\frac{\partial\phi}{\partial x} = -\sqrt{\frac{1}{2}\left(\frac{Q}{xk'}\right)} \qquad (46)$$

Equation (46) is plotted in Figure 11.

Conclusions—Theoretical solutions for the salt encroachment problem, assuming salt and fresh water to be immiscible, have been derived for six cases. The two basic cases are shown in Figures 1 and 6. Two special cases occur for aquifers infinite to the left in Figures 1 and 6 and the two final cases are for aquifers of infinite depth.

The numerical results for the infinite-depth aquifers (Fig. 11) represent extreme inclinations of the outflow face and indicate the limits of the location of the interface for intermediate cases.

Additional numerical results for particular cases of aquifers of finite depth and length in which the interface extends just to the upstream edge of the formation are shown in Figures 4 and 5.

In addition to the location of the interface, the numerical results also include distribution of velocities u_{AF} along the upper boundary of the aquifer and, for the finite formations, velocities u_{EF} along the vertical entrance to the aquifer. Also of interest is the location of y_0 when compared with the actual location of the interface. It is again noted that y_0 would give the location of the interface if lines of constant piezometric head were vertical instead of concave toward the sea. For an unconfined aquifer y_0 would be the location of the interface as estimated by the Ghyben-Herzberg principle of balance of pressures and, as can be seen in Figures 4, 5, and 11, this type of estimate usually indicates considerably more encroachment than actually occurs.

The results make available a rational theory which can be compared with field measurements and model tests. In future studies an attempt will be made to superimpose the effects of diffusion to obtain a more realistic representation of field conditions.

Acknowledgments—The writer wishes to thank Richard Skalak of Columbia University for constant encouragement and advice and for reviewing the manuscript. Also, thanks are due Emmett Laursen of Michigan State University for helpful discussions and suggestions. The study was suggested by discussions with R. R. Bennett and H. H. Cooper, Jr. of the U. S. Geological Survey, and their cooperation is gratefully acknowledged.

References

Glover, R. E., The pattern of fresh-water flow in a coastal aquifer, *J. Geophys. Research, 64,* 457–459, 1959.

Hamel, G., Über Grundwasserströmung, *Z. angew. Math. u. Mech., 14,* 130–157, 1934.

Hamel, G., and E. Günther, Numerische Durchrechnung zu der Abhandlung über Grundwasserströmung, *Z. angew. Math. u. Mech., 15,* 255–265, 1935.

Hayashi, K., *Tafeln der Besselschen, Theta, Kugel, und anderer Funktionen,* Springer, Berlin, 125 pp., 1930.

Jahnke, Eugene, and Fritz Emde, *Tables of Functions,* Dover Publications, New York, 304 pp., 1945.

Kozeny, Joseph, *Hydraulik,* Springer, Wien, 588 pp., 1953.

Muskat, Morris, The Seepage of Water through Dams with Vertical Faces, *Physics, 6,* 402–415, 1935.

Muskat, Morris, *The Flow of Homogeneous Fluids through Porous Media,* McGraw-Hill, New York, 763 pp., 1937.

Nehari, Zeev, *Conformal Mapping,* McGraw-Hill, New York, 396 pp., 1952.

Polubarinova-Kochina, P. Y., On the unsteady motion of ground water in two layers of different density (in Russian), *Izvest. Akad. Nauk SSSR, Otdel Tekh. Nauk, No. 6,* 1940.

(Manuscript received July 16, 1959; presented at the Fortieth Annual Meeting, Washington, D. C., May 5, 1959.)

Journal of Geophysical Research Volume 64, No. 11 November, 1959

Analysis of Data From Pumping Wells Near a River

Mahdi S. Hantush

New Mexico Institute of Mining and Technology, Socorro, New Mexico[1]

Abstract—Methods are outlined for determining the hydrologic characteristics of an aquifer draining or drained by a stream; specifically, for determining the transmissibility and storage coefficient of the acquifer and also the effective distance from a pumped well to a line in the stream bed where the water is entering or leaving the aquifer. The procedure is based on the theory of nonsteady flow toward a well steadily discharging from an infinite acquifer hydraulically connected to a fairly straight and long stretch of a stream bed. Application of these methods is illustrated by treating data from the Ingalls area in Kansas.

Introduction

In quantitative investigations of ground-water resources it is important to know the field values of the so-called formation constants. These constants describe the hydraulic properties of the aquifers. Knowledge of these parameters makes it possible to predict the hydraulic behavior of the ground-water system.

In recent years, emphasis has been given to the installation of wells in sands and gravel deposits along perennial streams for the purpose of inducing infiltration from these streams to the water-bearing materials underlying them, or for the purpose of salvaging natural flow which would otherwise have been discharged to the streams. Accordingly the yield of the aquifers is increased by the amount of induced recharge or the amount of salvage, as the case may be. Another object of such installation may be to deter salt encroachment in aquifers in the coastal plains by recharging wells, in which case the effect of the presence of the stream on the ground-water flow is replaced by that of the sea. The flow toward a pumping (or a recharging) center in such a ground-water system, where the stream or the sea represents a line of constant head, does not depend on the formation constants only; it depends also on the effective distance between the center of pumping and the line of constant head. The determination of this effective distance is important in the investigation of such ground-water systems. As shown by experience, serious errors may result if the values of the effective distance are based on general assumptions, such as the distance from the center of production wells to the bank or to the midline of the stream channel.

Based on the steady-state distribution of drawdown around a well operating near a river, a method for determining the effective distance and the coefficient of transmissibility has been described by *Rorabaugh* [1956]. Because equilibrium conditions are assumed (a condition which may not be attained during ordinary periods of pumping), the storage coefficient is not determined by this method.

Kazmann [1946] has proposed a matching-curve method for obtaining the effective distance. In 1948, he described a second method for the same objective [*Kazmann,* 1948]. Both of these methods require preknowledge of the formation constants, which Kazmann obtained by applying Theis' graphical method to the early part of the data. Such a determination may give erroneous results, for two reasons: first, because the recharge from the stream may affect the flow pattern even though its influence may not be apparent on the time-drawdown curve, and, second, because of the probable changes of the formation coefficients, especially during the early period of pumping. This is especially true in water-table aquifers.

The purpose of this paper is to outline methods by which the effective distance to the line of constant head, as well as the formation

[1] On leave from the College of Engineering, University of Baghdad, Iraq.

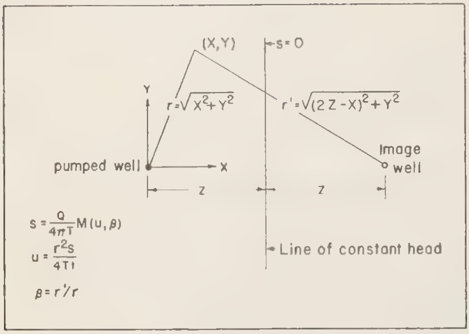


FIG. 1—Diagrammatic representation of a steady well near a line of constant head.

coefficients, may be obtained by using data from the period during which the effect of the line of constant head on the flow system takes place. The formation coefficients tend more or less to attain their constant values as the time of pumping increases. Thus, the avoidance of data collected during the early period of pumping will give truer values for the formation coefficients.

THEORY

The drawdown at any point in the vicinity of a well that is steadily discharging from a homogeneous, uniform, isotropic aquifer which is hydraulically connected to an infinitely long line of constant head is given by *Jacob* [1950]:

$$s = (Q/4\pi T)M(u, \beta) \qquad (1)$$

where

$$M(u, \beta) = W(u) - W(\beta^2 u)$$

$$u = \alpha/t \qquad (2)$$

$$\alpha = r^2 S/4T \qquad (3)$$

$$\beta = r'/r \qquad (4)$$

and where s is the drawdown of the piezometric surface at any time t since the start of pumping and at any distance $r = \sqrt{x^2 + y^2}$ from the well center; $r' = \sqrt{(2z - x)^2 + y^2}$ is the distance from the center of the image well to the point in question (Fig. 1); Q is uniform discharge; S and T are the storage and transmissibility coefficients respectively; W is the exponential integral or what is known as the well function; x and y are the coordinates of any point with respect to the center of the well, as shown in Figure 1; and z is the effective distance between the well center and the line of constant head.

Basically equation (1) is for artesian conditions; it can be used, however, for water-table conditions with fair assurance if s is replaced by s' (that is, by $s - s^2/2h_o$), provided that $s/h_o \leq 0.25$ and that the storage coefficient remains fairly constant [*Jacob*, 1950], h_o being the natural or initial depth of flow and T being the transmissibility coefficient corresponding to this depth.

The relation between β, x, r, and z is given by

$$4z^2 - 4xz - r^2(\beta^2 - 1) = 0 \qquad (5)$$

Equation (1) can be expanded in a convergent series as

$$\begin{aligned} s &= (Q/4\pi T) \\ &\quad \cdot\left[2 \ln \beta + \sum_{n=1}^{\infty} (-)^n(\beta^{2n} - 1)\alpha^n/n\cdot n!\right] \\ &= (Q/4\pi T)[2 \ln \beta - (\beta^2 - 1)\alpha/t \\ &\quad + (\beta^4 - 1)\alpha^2/2\cdot 2!t^2 - \cdots] \qquad (6) \end{aligned}$$

For small values of $(\beta^2 u)$, that is, large relative values of time $(\beta^2 u \leq 0.10)$, (6) may be approximated by

$$s \approx (Q/4\pi T)[2 \ln \beta - (\beta^2 - 1)\alpha/t + (\beta^4 - 1)\alpha^2/2\cdot 2!t^2] \qquad (7)$$

and, for values of $\beta^2 u < 0.05$, by

$$s \approx (Q/4\pi T)[2 \ln \beta - (\beta^2 - 1)\alpha/t] \qquad (8)$$

Properties of the solution—Some properties of (1) follow.

A. The theoretical graph of the drawdown s against the time t on semilogarithmic paper, with t plotted in the logarithmic scale (Fig. 2), has the following properties:

(a) The curve has an inflection point at which the following relation holds

$$u_i = \alpha/t_i = \frac{2 \ln \beta}{\beta^2 - 1} \qquad (9)$$

where the subscript i pertains to values of the variables at the inflection point. This relation is

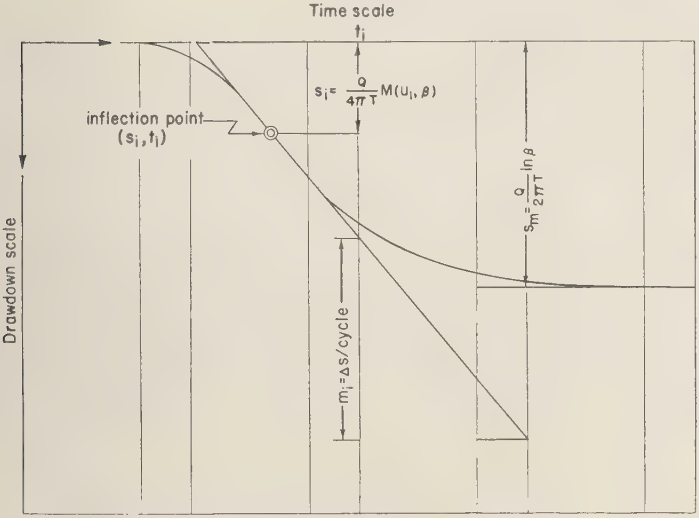

Time scale
t_i
$s_i = \frac{Q}{4\pi T} M(u_i, \beta)$
inflection point (s_i, t_i)
Drawdown scale
$m_i = \Delta s/\text{cycle}$
$s_m = \frac{Q}{2\pi T} \ln \beta$

FIG. 2—Time-drawdown variation due to a steady well near a line of constand head.

btained by equating to zero the second derivaive of s with respect to $\ln t$ and solving for u.

(b) The slope of the curve m at any point is iven by

$$m = (2.3Q/4\pi T)[e^{-u} - e^{-\beta^2 u}] \qquad (10)$$

Equation (10) is obtained by differentiating s n (1) with respect to $\log_{10} t$.

(c) The slope of the curve m_i at the inflection oint is

$$m_i = (2.3Q/4\pi T)[e^{-u_i} - e^{-\beta^2 u_i}] \qquad (11)$$

which is obtained by substituting u_i for u in (10).

(d) The drawdown s_i at the inflection point s obtained by replacing u by u_i in (1); that is,

$$s_i = (Q/4\pi T) M(u_i, \beta) \qquad (12)$$

(e) For increasing values of t, the curve approaches the maximum drawdown (steady state) s_m; namely, it approaches the value given by

$$s_m = (Q/2\pi T) \ln \beta \qquad (13)$$

which is obtained by letting t become infinite in (1) or (6).

(f) The ratio (s_m/m_i) depends on the value of β only, so that

$$s_m/m_i = \frac{2 \log_{10} \beta}{e^{-u_i} - e^{-\beta^2 u_i}} = f(\beta) \qquad (14)$$

where u_i is given by (9).

B. The theoretical curve of the drawdown s against the reciprocal of time $(1/t)$ has (Fig. 3) the following properties:

(a) It approaches asymptotically the $(1/t)$-axis.

(b) It is approximately the parabola given by (7) in the range where $\beta^2 u < 0.10$; that is, the parabola given by

$$s = s_m - m_t/t + c/t^2 \qquad (15)$$

where s_m, m_t, and c represent the corresponding coefficients of $(1/t)$ given by (7).

(c) It is approximately the straight line given by (8) in the range $\beta^2 u < 0.05$, or

$$s = s_m - m_t/t \qquad (16)$$

(d) It has an s-intercept equal to the maximum drawdown s_m.

(e) Its slope at the s-intercept is equal to $(-m_t)$, which is equal to

$$m_t = (Q/4\pi T)(\beta^2 - 1)\alpha \qquad (17)$$

Equation (17) is obtained by differentiating (6) with respect to $(1/t)$ and letting $(1/t)$ equal zero.

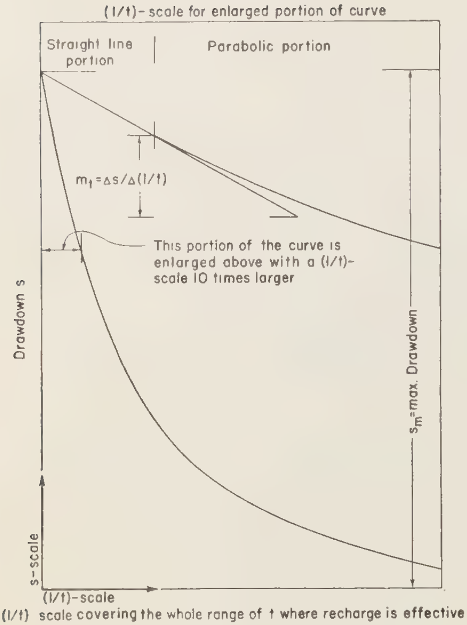


Fig. 3—(1/t)-drawdown variation due to a steady well near a line of constant head.

With s_m as the s-intercept and $(-m_t)$ as the slope, the equation of the tangent line to the curve at the intercept point is obviously given by (16). The tangent line cuts the $(1/t)$-axis at a value of t given by

$$t = m_t/s_m \tag{18}$$

Since s_m and m_t are given by (13) and (17) respectively, equation (18) can be written as

$$\alpha/t = u = \frac{2 \ln \beta}{\beta^2 - 1} \tag{19}$$

If one compares (9) and (19), it is readily seen that u of (19) is equal to u_i, and therefore the value of t as obtained by (18) is the value of the time t_i at which the inflection point of the time-drawdown semilogarithmic curve occurs. Thus, the value of t_i can be obtained from (18).

(f) The expression relating the values of s_m, m_t, and c of (15) to the value of β is given by

$$F(\beta) = \frac{\beta^2 + 1}{\beta^2 - 1} \log_{10}\beta = 0.87cs_m/m_t^{\,2} \tag{20}$$

Equation (20) is obtained by eliminating α and $Q/4\pi T$ from the expressions relating these parameters to s_m, m_t, and c.

C. Properties of equation (17). If one replaces α and β by their respective values in (3) and (4) and substitutes in (17), the resulting expression, after simplification, is

$$m_t = (QSz/4\pi T^2)(z - x) \tag{21}$$

Thus, a plot of m_t versus x is a straight line whose slope $(-m_x)$ is given by

$$m_x = QSz/4\pi T^2 \tag{22}$$

and whose x-intercept x_0 is equal to z, the effective distance to the line of constant head.

D. Properties of equation (13). A semilogarithmic curve of s_m versus β (β being on the logarithmic scale) is a straight line passing through the point $s_m = 0$, $\beta = 1$ and having a slope equal to

$$m_m = 2.3Q/2\pi T \tag{23}$$

Tables of functions—In order to apply the above theory for determining the different parameters, tables of the functions $M(u_i, \beta)$, $f(\beta)$, u_i, and $F(\beta)$are necessary. The M function can, of course, be computed by the use of tables of the well function W, which has already been tabulated [*Wenzel*, 1942; *Wisler and Brater*, 1951]. These functions are given in Tables 1 and 2 for values of β that are of practical use. For values of $\beta > 100$, the functions may be approximated by

$$f(\beta) \approx 2 \log_{10} \beta \tag{24}$$

$$F(\beta) \approx \log_{10} \beta \tag{25}$$

$$M(u_i, \beta) \approx 2.3 \log_{10} (0.562/u_i) \tag{26}$$

The least-squares method—The unknown coefficients of (15) can be found by the use of the method of least squares, provided the observational data fall within the period during which the drawdown is approximated by (15). Since there are 3 unknowns in (15), there must be 3 normal equations for its solution. These are

$$\begin{aligned} \sum s &= ns_m - m_t \sum (1/t) + c \sum (1/t)^2 \\ \cdot \sum (1/t)s &= s_m \sum (1/t) \\ &\quad - m_t \sum (1/t)^2 + c \sum (1/t)^3 \\ \cdot \sum (1/t)^2 s &= s_m \sum (1/t)^2 \\ &\quad - m_t \sum (1/t)^3 + c \sum (1/t)^4 \end{aligned} \tag{27}$$

where n is the number of points used.

Table 1—*Values of the functions* u_i, $M(u_i, \beta)$, *and* $f(\beta)$

β	u_i	$M(u_i, \beta)$	$f(\beta)$	β	u_i	$M(u_i, \beta)$	$f(\beta)$	β	u_i	$M(u_i, \beta)$	$f(\beta)$	β	u_i	$M(u_i, \beta)$	$f(\beta)$
1.0	1.000	0.000	1.179	5.0	0.134	1.553	1.667	10	0.0466	2.534	2.115	35	0.00582	4.576	3.109
1.1	0.909	0.070	1.183	5.2	0.127	1.604	1.688	11	0.0400	2.680	2.188	36	0.00554	4.624	3.134
1.2	0.830	0.135	1.188	5.4	0.120	1.653	1.710	12	0.0348	2.815	2.251	37	0.00528	4.671	3.155
1.3	0.761	0.195	1.194	5.6	0.114	1.703	1.731	13	0.0306	2.940	2.312	38	0.00505	4.717	3.178
1.4	0.702	0.252	1.203	5.8	0.108	1.750	1.752	14	0.0271	3.057	2.367	39	0.00483	4.761	3.199
1.5	0.649	0.306	1.214	6.0	0.102	1.796	1.770	15	0.0241	3.172	2.423	40	0.00462	4.805	3.221
1.6	0.603	0.357	1.223	6.2	0.0976	1.840	1.794	16	0.0218	3.271	2.472	41	0.00443	4.847	3.242
1.7	0.562	0.407	1.235	6.4	0.0930	1.988	1.814	17	0.0203	3.342	2.520	42	0.00424	4.889	3.262
1.8	0.525	0.456	1.247	6.6	0.0888	1.927	1.833	18	0.0179	3.462	2.564	43	0.00407	4.930	3.282
1.9	0.492	0.502	1.262	6.8	0.0848	1.969	1.852	19	0.0164	3.551	2.609	44	0.00391	4.969	3.301
2.0	0.462	0.548	1.273	7.0	0.0812	2.010	1.871	20	0.0150	3.637	2.647	45	0.00376	5.008	3.321
2.2	0.411	0.635	1.301	7.2	0.0777	2.050	1.889	21	0.0138	3.716	2.687	46	0.00362	5.046	3.339
2.4	0.368	0.717	1.329	7.4	0.0745	2.089	1.908	22	0.0128	3.793	2.725	47	0.00349	5.084	3.357
2.6	0.332	0.796	1.357	7.6	0.0715	2.127	1.925	23	0.0119	3.867	2.761	48	0.00336	5.120	3.375
2.8	0.301	0.872	1.385	7.8	0.0687	2.165	1.943	24	0.0111	3.938	2.796	49	0.00325	5.156	3.393
3.0	0.275	0.945	1.413	8.0	0.0661	2.202	1.960	25	0.0103	4.007	2.837	50	0.00313	5.191	3.410
3.2	0.252	1.016	1.435	8.2	0.0636	2.238	1.977	26	0.00966	4.072	2.862	55	0.00265	5.358	3.491
3.4	0.232	1.083	1.467	8.4	0.0613	2.273	1.994	27	0.00906	4.135	2.893	60	0.00228	5.510	3.565
3.6	0.214	1.149	1.493	8.6	0.0590	2.308	2.010	28	0.00852	4.196	2.923	65	0.00198	5.650	3.634
3.8	0.199	1.212	1.500	8.8	0.0570	2.342	2.026	29	0.00803	4.256	2.952	70	0.00174	5.781	3.697
4.0	0.185	1.273	1.545	9.0	0.0550	2.376	2.041	30	0.00757	4.313	2.980	75	0.00154	5.903	3.757
4.2	0.173	1.333	1.571	9.2	0.0531	2.408	2.057	31	0.00716	4.369	3.008	80	0.00137	6.017	3.812
4.4	0.162	1.390	1.597	9.4	0.0513	2.441	2.072	32	0.00678	4.423	3.034	85	0.00123	6.124	3.864
4.6	0.152	1.447	1.619	9.6	0.0497	2.472	2.087	33	0.00643	4.475	3.059	90	0.00111	6.226	3.913
4.8	0.142	1.500	1.642	9.8	0.0481	2.503	2.102	34	0.00611	4.526	3.085	95	0.00102	6.311	3.960
5.0	0.134	1.553	1.667	10.0	0.0466	2.534	2.115	35	0.00582	4.576	3.109	100	0.00092	6.412	4.004

Application

Depending on the number of observation wells available, one of the following procedures may be followed in calculating the formation coefficients. It is assumed, of course, that the period of pumping is long enough so that the maximum drawdown can be extrapolated.

Each of the three procedures described below requires an estimate of the maximum drawdown s_m. An approximate value of s_m can be obtained (a desirable but not necessary step) by the method of least squares in fitting the parabola of (15) to the points $(s, 1/t)$ collected during the period of recharge which is indicated by the semilogarithmic plot of time against drawdown. Because of the uncertainty of determining the time after which the $(1/t)$-drawdown curve becomes parabolic, the values of the coefficients c and m_i may be in great error. The value of s_m, on the other hand, although not accurately determined, is close enough to be of value in constructing the time-drawdown curve.

One observation well—For one observation well, there are two procedures that may be followed.

First procedure: On semilogarithmic paper, plot time versus drawdown (time in any convenient unit, usually in minutes, being on the logarithmic scale). If the approximate value of s_m is obtained by the method of least squares, this value should be indicated on the graph also. Construct the time-drawdown curve through the plotted points. Extrapolate the value of the maximum drawdown s_m. Construct the tangent line to the curve at the inflection point and measure its slope m_i. Because the position of the inflection point is not known, this tangent line is usually approximated by the straight portion of the curve as extrapolated from the trend of the curve in the period of effective recharge. Compute $f(\beta)$, which is the ratio s_m/m_i, and obtain the corresponding value of β, using Table 1. Compute T, using equation (13). Obtain the values of u_i and $M(u_i, \beta)$, using Table 1; then obtain the corresponding value of s_i, using (12). Knowing s_i, locate the inflection point on the curve and read t_i. Compute S from (9); that is, from the relation $S = 4Tt_iu_i/r^2$. Compute z from (5).

Theoretically, when the calculated parameters are used in conjunction with equation (1) and with assigned values of t, the calculated drawdowns should fall on the observed curve. This, however, is usually not true for the early portion of the observed curve, because generally the formation coefficients vary in the early stages of pumping. During the latter part of the test, or as time increases, the coefficients usually tend to attain their uniform values. Direct computation in equation (1) for drawdowns in the latter part of the test should compare fairly well with observed values. Sometimes the

TABLE 2—*Values of the function* $F(\beta) = \beta^2 + 1)/(\beta^2 - 1) \log_{10} \beta$

β	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
1.0	0.434	0.436	0.439	0.444	0.451	0.458	0.466	0.474	0.483	0.492	0.502
2.0	0.502	0.511	0.521	0.530	0.540	0.550	0.560	0.568	0.578	0.587	0.600
3.0	0.600	0.605	0.614	0.623	0.632	0.641	0.650	0.658	0.666	0.674	0.682
4.0	0.682	0.690	0.698	0.706	0.713	0.720	0.728	0.736	0.743	0.750	0.757
5.0	0.757	0.764	0.771	0.778	0.784	0.791	0.797	0.804	0.810	0.816	0.823
6.0	0.823	0.829	0.835	0.841	0.846	0.852	0.858	0.863	0.869	0.874	0.880
7.0	0.880	0.885	0.891	0.896	0.902	0.907	0.912	0.917	0.922	0.927	0.932
8.0	0.932	0.936	0.941	0.946	0.951	0.955	0.960	0.964	0.969	0.973	0.978
9.0	0.978	0.982	0.987	0.991	0.996	1.000	1.004	1.008	1.012	1.016	1.020
10.0	1.020	1.024	1.028	1.032	1.036	1.040	1.044	1.047	1.051	1.055	1.059
11.0	1.059	1.062	1.066	1.069	1.073	1.077	1.081	1.084	1.087	1.090	1.094
12.0	1.094	1.097	1.101	1.104	1.108	1.111	1.114	1.117	1.121	1.124	1.127
13.0	1.127	1.130	1.133	1.136	1.140	1.143	1.146	1.149	1.152	1.155	1.158
14.0	1.158	1.161	1.164	1.167	1.170	1.172	1.175	1.178	1.181	1.184	1.187
	0	1	2	3	4	5	6	7	8	9	10
10	1.020	1.059	1.094	1.127	1.158	1.187	1.214	1.239	1.263	1.286	1.308
20	1.308	1.328	1.348	1.367	1.385	1.402	1.419	1.437	1.451	1.466	1.480
30	1.480	1.494	1.508	1.521	1.534	1.547	1.559	1.570	1.582	1.593	1.604
40	1.604	1.615	1.625	1.635	1.645	1.655	1.664	1.674	1.683	1.692	1.700
50	1.700	1.709	1.717	1.725	1.734	1.741	1.749	1.757	1.764	1.772	1.779
60	1.779	1.786	1.793	1.800	1.807	1.814	1.820	1.827	1.833	1.840	1.846
70	1.846	1.850	1.858	1.864	1.870	1.876	1.881	1.887	1.892	1.898	1.904
80	1.904	1.909	1.914	1.919	1.925	1.930	1.935	1.940	1.945	1.950	1.955
90	1.955	1.959	1.964	1.969	1.974	1.978	1.983	1.987	1.992	1.996	2.000

extrapolated value of s_m (as obtained above) is either over- or underestimated, and/or the slope of the curve at the inflection point is drawn either flatter or steeper than necessary. In such cases, the calculated drawdowns necessarily will deviate from the observed curve. The next step is, therefore, to adjust the extrapolation of s_m, and/or the slope of the curve at the inflection point. The amount of adjustment is inferred from the amount and direction of the deviation of the calculated values of drawdown from those observed. Then the above steps are repeated. With a little experience, only one trial may be needed to arrive at a satisfactory answer.

Second procedure. It should be emphasized before introducing the second procedure that it will not yield good results unless care and precision in data collection have been exercised, and unless the field conditions approach very closely the assumptions made in the development of the theory. Although the method may be of interest to hydrologists, it may be more applicable to problems in heat conduction, where precision in data collection may be achieved and where the conditions imposed by the theoretical treatment may obtain.

The procedure is as follows: Carry out step 1 of the first procedure. Select the points which occur in the range where the effect of recharge is apparent. Eliminate from these points all that deviate greatly from the general trend of the curve. Using the remaining points, apply the method of least squares to determine the coefficients s_m, m_i, and c of (15), taking s and $(1/t)$ as the dependent and the independent variables, respectively. Compute $F(\beta)$ from (20) and find the corresponding value of β from Table 2. Compute T from (13) and then S from (17). Compute z from (5).

Two or more observation wells—The procedure outlined above may, of course, be applied to each well independently. The coefficients thus computed are then weighted to obtain their mean values in the vicinity of the well field. A procedure that would average the formation coefficients may be employed, however, if there are at least two observation wells. The procedure follows: (1) Construct the time-drawdown curve on semilogarithmic paper, as discussed above. (2) Construct the $(1/t)$-drawdown curve, using all the observational data that fall within the range where the recharge is effective, as indicated by the curve of step 1. The reciprocal of time $1/t$ (time in any convenient units, usually hours or days) is plotted on a scale such that all the points are included. In constructing this curve, advantage must be taken of the properties of the theoretical curve. (3) Using a larger scale for $1/t$ than that used in step 2, construct the curve for most of the latter part of the data, making use of the general trend of the curve constructed in step 2. (4) Construct the tangent to the curve at the s-intercept and measure its slope $(-m_t)$. Measure also the s-intercept. (5) Knowing m_t and x (position of observation well), plot m_t versus x on uniform scales and construct the best-fit straight line through the plotted points. The equation of this line is given by (21). (6) Measure the slope of this line $(-m_x)$ and its x-intercept. This intercept is the value of the effective distance to the line of recharge; that is, the value of z. It should be remarked that the plotted points of m_t versus x may be widely scattered. In general, this is due to the nonuniformity of the formation coefficients. It may also be due, however, to errors in estimating the values of s_m and m_t, both of which are subject to over- or underestimation. In such cases, the next step is to adjust the values of s_m or m_t, or both, by reexamining the fit of the $(1/t)$-drawdown curve. The amount of adjustment is inferred from the amount of scatter in the already plotted points of m_t versus x. Several trials may be required before a satisfactory fit is obtained. With a little experience, one trial may be sufficient. If, after these adjustments are made, the scatter of the points persists, this indicates that the formation constants vary widely in the field of pumping, in which case an attempt to find average values for these constants is superfluous. (7) Compute β from (4). (8) On semilogarithmic paper, plot s_m versus β (β being on the logarithmic scale), construct the best-fit straight line through the point ($s_m = 0, \beta = 1$) and the plotted points, and measure the slope m_m of this line. If a straight line cannot be made to fit fairly well, a nonuniform transmissibility in the field of pumping is indicated. In such cases, a minimum value of T for the purpose of safe design can be obtained by constructing a line through the origin which envelops all plotted points, instead of using the average of plotted points. (9) Compute T from the relation $T = 2.3Q/2\pi m_m$. (10) Compute S from (22).

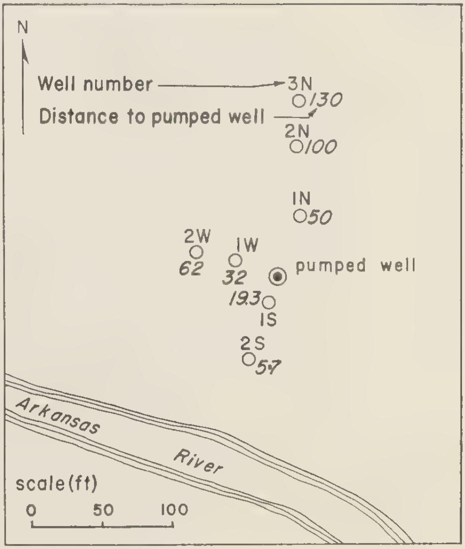


FIG. 4—Location of wells at aquifer test site.

Examples—Aquifer test data obtained at the Norbert Irsik site near the Arkansas River, in the Ingalls area, Kansas [*Stramel and others*, 1958], are here used to illustrate the application of the methods presented above. The location of the wells used in the pumping test is shown in Figure 4. The pumped well, 16 inches in diameter, 29 feet deep, and screened with perforated pipe, is 135 feet from the Arkansas River. The well completely penetrates the alluvial deposit. The thickness of the saturated material prior to the test was 22 feet. The observation wells partially penetrate the aquifer. The well was pumped at an average rate of 1.55 ft^3/sec. Because the flow is unconfined (water-table conditions), the observed drawdowns are adjusted by a subtractive factor equal to $s^2/2h_0$, or $s^2/44$, to give values of the adjusted drawdown s'. The semilogarithmic s'-time curves for the seven observation wells are shown in Figures 5 and 6.

As pointed out previously, the probable

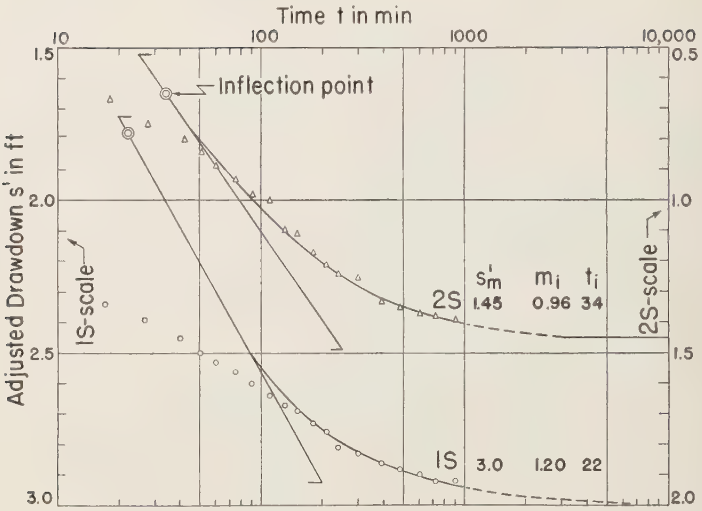


FIG. 5—Time-drawdown variation in wells 1*S* and 2*S*.

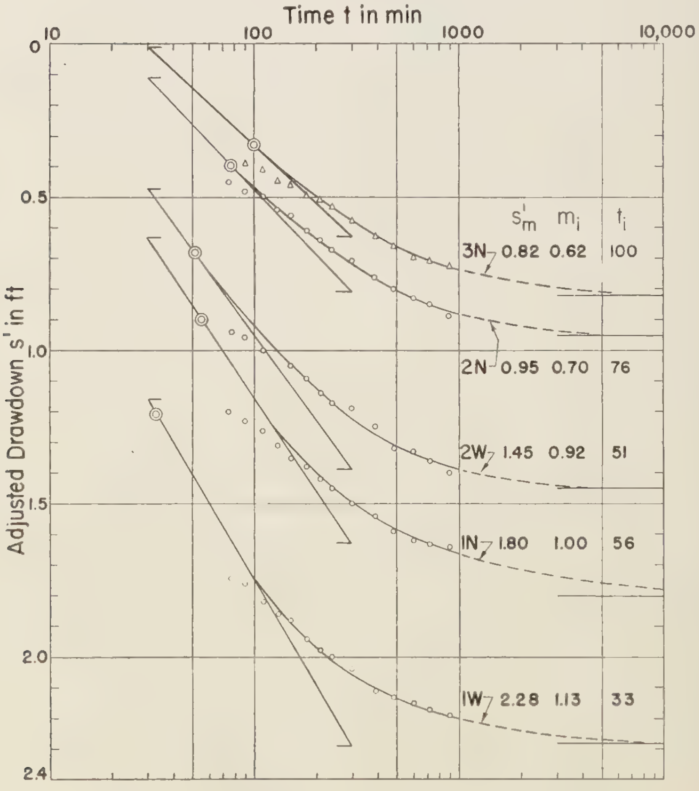


FIG. 6—Time-drawdown variation in the observation wells.

TABLE 3—*Computation of formation parameters using first method*

Well No.	r (ft)	s_m (ft)	m_i	$f(\beta)$	β	$\log_{10}\beta$	u_i	$M(u_i, \beta)$	$(Q/4\pi T)$ (ft)	s_i (ft)	t_i (min)	z (ft)	T (ft^2/sec)	S
1S	19.3	3.00	1.20	2.50	16.7	1.22	0.021	3.32	0.535	1.78	22	165	0.232	0.067
1W	32.0	2.28	1.13	2.02	8.8	0.96	0.057	2.33	0.518	1.20	33	141	0.240	0.105
1N	50.0	1.80	1.00	1.80	6.3	0.80	0.096	1.83	0.490	0.90	56	132	0.253	0.131
2W	62.0	1.45	0.92	1.58	4.3	0.63	0.167	1.35	0.517	0.70	52	130	0.240	0.128
2S	57.0	1.45	0.96	1.51	3.8	0.58	0.20	1.20	0.543	0.65	34	137	0.228	0.114
2N	100.0	0.95	0.70	1.36	2.6	0.42	0.34	0.80	0.492	0.39	76	80[a]	0.252	0.156
3N	130.0	0.82	0.62	1.32	2.3	0.36	0.39	0.67	0.495	0.33	100	85[a]	0.250	0.138
Average Values												141*	0.242	0.116**

a Indicating additional source of recharge (may be due to increase in formation thickness).

* Arithmatic average excluding wells 2N and 3N

** Logarithmic average

changes of the formation coefficients with time during early periods of pumping (especially true if the flow is unconfined) result in an entirely different time-drawdown variation from that obtained when the coefficients remain constant with time. As time progresses, the coefficients tend to attain their uniform values, and the time-drawdown variation approximates the trend indicated by theory. Thus, in constructing the semilogarithmic time-drawdown curve, the points collected prior to the period of recharge are neglected.

An approximate value of the maximum drawdown is obtained by the method of least squares, fitting the parabola of (15) through the points $(s', 1/t)$ collected during the period of recharge. This value of s'_m, together with the points collected during the recharge period, is used to construct the best-fit semilogarithmic time-drawdown curves.

First case, first procedure. The schedule for computing the parameters is given in Table 3. The computation procedure for well 1*S* follows: From Figure 5, s'_m is 3.00 ft. The curve passing through the points in the period where recharge has occurred is extended to the left until it has a straight-line trend. This portion of the curve is taken as its tangent at the inflection point. Its slope m_i (conveniently taken as $\Delta\ s'$ /cycle) is measured as 1.20. Then $f(\beta) = s_m/s_i = 2.50$. For this value of $f(\beta)$, Table 1 gives $\beta = 16.7$. With this value of β, the values of $M(u_i, \beta)$, u_i, and $\log_{10} \beta$ are respectively (Table 1) 3.32, 0.021, and 1.22. Equation (13) gives $Q/4\pi T = 0.535$; therefore, $T = 0.232$ ft^2/sec or 151,000 gpd/ft. From (12), $s_i = 1.78$ ft; then from Figure 5, $t_i = 22$ min $= 1{,}320$ sec. The storage coefficient is then computed from $S = 4Tt_iu_i/r^2 = 4(0.232)\ (1{,}320)(0.021/19.3^2) = 0.067$. For $\beta = 16.7$ and $x = 19.3$, (5) gives $z = 165$ ft. The computation is carried out by slide rule.

Figure 5 shows the curves for wells 1*S* and 2*S*, in which the points collected during the early period of pumping are retained to show how great the difference will be between the early part of the observed curve and that obtained by assuming constant formation coefficients during the later part of the test. Had all the points been given the same weight, the application of the theory on well 1*S* would have given a transmissibility coefficient approximately three times that obtained above, and would have given a value of z, the effective distance to line of constant head, approximately equal to 5000 ft, values not compatible with the geological information at hand.

The values of z obtained by analyzing data from wells 2*N* and 3*N* indicate that the amount of induced recharge could not have been obtained from the Arkansas River only. Local recharge must have taken place. This is most probably due to an increase in aquifer thickness in the area.

First case, second procedure. As mentioned previously, this method gives reasonable results only if the data used fall in the period when the curve can be approximated safely by the

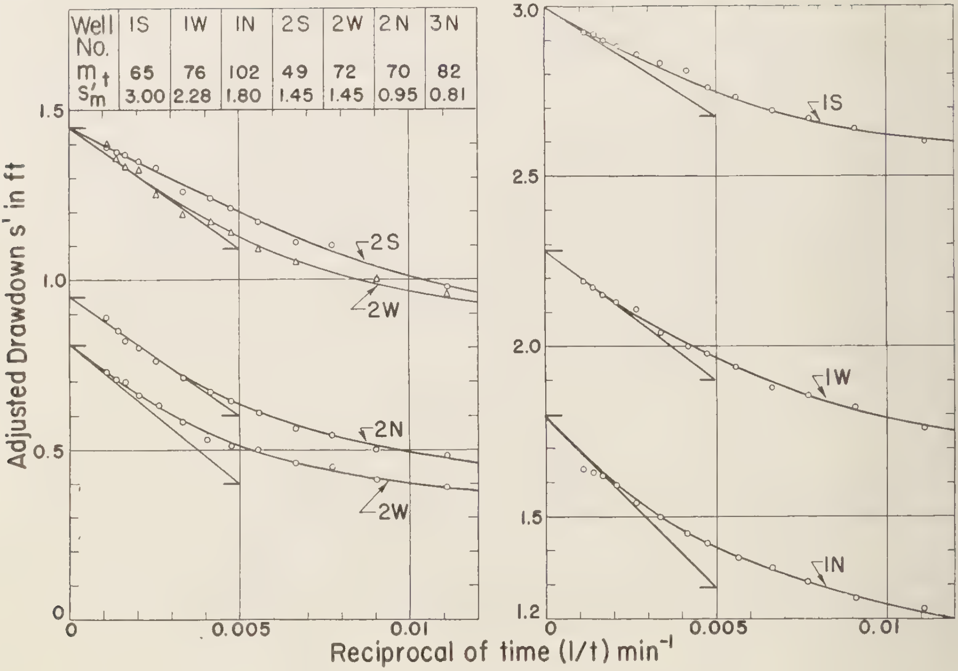


FIG. 7—(1/*t*)-drawdown variation in the observation wells.

parabola of (15). If some of the selected points from which the equation of the parabola is to be determined do not fall in the region where the approximation can be made, the resulting equation, although parabolic, may deviate greatly from the equation actually sought. The largest difference is in the value of the coefficient of the third term, the difference being somewhat less in that of the second. The value of the first term is least affected. A better representation of the actual curve may be obtained by taking the points collected during the period when recharge has occurred and fitting them to a fourth-degree equation via (6). In this case,

TABLE 4—*Computation of formation parameters using second method*

Well No.	s'm (ft)	m_t (ft-min)	r (ft)	x (ft)	z* (ft)	r' (ft)	m_x* (min)	β	m_m*	T* (ft^2/sec)	S
1S	3.00	65	19.3	19.3	140	261	0.53	13.6	2.32	0.244	0.11
1W	2.28	76	32.0	0	"	280	"	8.8			
1N	1.80	102	50.0	-50.0	"	330	"	6.6			
2W	1.45	72	62.0	0	"	286	"	4.6			
2S	1.45	49	57.0	57.0	"	223	"	3.9			
2N	0.95	70	100.0	-97**	80^a	260	0.40	2.6			
3N	0.82	82	130.0	-127**	80^a	290	"	2.2			

* Average value for the group indicated

** Estimated value

a See note in table 3

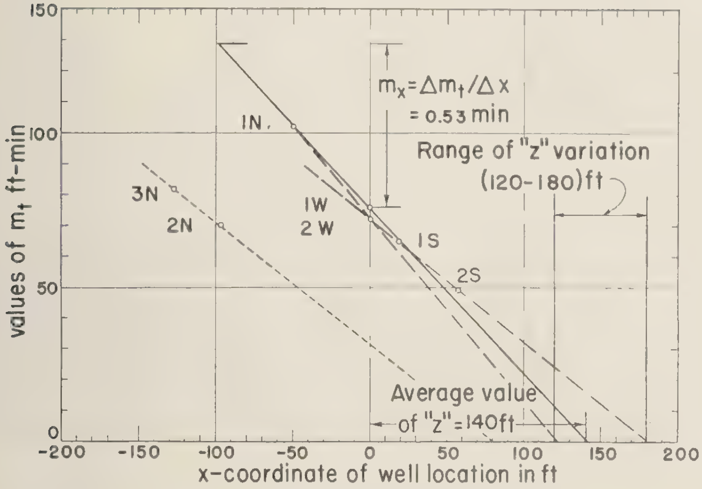


FIG. 8—Variation of m_t with well location.

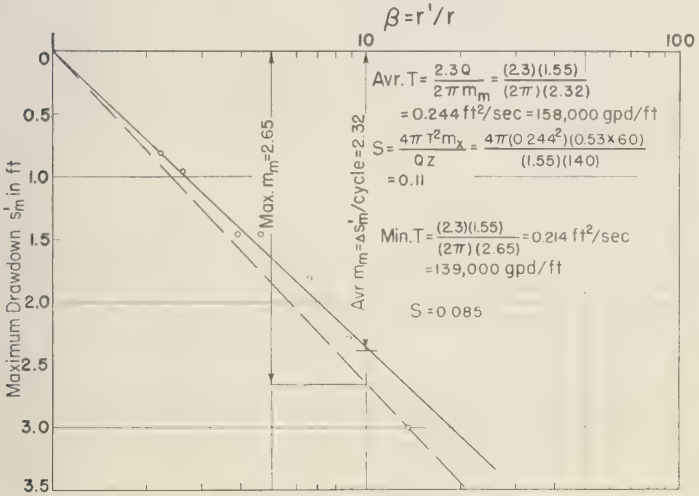


FIG. 9—Variation of s'_m with β.

however, the method loses its advantage of brevity, because the computations become lengthy and more laborious, and the chance of computational error increases unless a mechanical computor is used.

When this method is applied to the data under discussion, the results are in conformity neither with those obtained by the first procedure nor with those obtained by the second method. This is due, apparently, to the fact that the points collected are not in the range where the curve can be approximated by the parabola of (15); that is, the period of pumping was not long enough to insure a sufficient number of points in the range where the curve is parabolic. For the purpose of illustrating the procedure, however, the analysis of data from well 1S is presented here, although the results obtained are not dependable. The points used in the analysis are those which fall within the period when recharge apparently has taken place. In the case of well 1S, the points are those for $t > 130$ minutes. The points which appear to deviate greatly from the general trend of the curve are disregarded.

If one applies the method of least squares to the remaining points in the equation

$$s' = s_m' - m_t/t + c/t^2$$

the coefficients s_m' m_t, and c are found to be 3 ft, 57.1 ft-min., and 1,910 ft-min.^{2} respectively. Using the relation $F(\beta) = 0.87\ s_m c/m_t^2$, $F(\beta) = 1.525$. Then, from Table 2 $\beta = 33.3$, and from (5), $z = 330$ ft. From $T = (2.3Q/2\pi s_m) \log_{10}\beta$, $T = 0.29$ ft^2/sec. And from $S = 4Tm_t/r^2 (\beta^2 - 1)\ (Q/4\pi T)$, $S = 4(0.29)\ (57.1 \times 60)/(19.3)^2\ (33.3^2-1)\ (0.32) = 0.03$.

Second case. Curves of $(1/t)$ versus drawdown are constructed for each observation well, as shown in Figure 7. Table 4 gives the schedule for computing the formation parameters. The tangents at the s'-intercept are drawn and their slopes are measured. Values of s_m' (s'-intercept) and of the slopes ($-m_t$) are then recorded. The values of m_t are plotted against the values of x, as shown in Figure 8. With the exception of the points of wells 2N and 3N, the points seem to approximate a straight line. Through these points, three straight lines are constructed: (1) The best-fit straight line (2) the line of steepest slope, and (3) the line of flattest slope. Their x-intercepts give, respectively, the average value of $z = 140$ ft, the minimum value of $z = 120$ ft, and the maximum value of $z = 180$ ft. The two points of 2N and 3N, taken together, give a value of $z = 80$ ft. (See previous discussion with respect to the behavior of these two wells.) The slopes ($-m_x$) of these lines are measured and recorded. Using the average value of $z = 140$ ft, the value of β for each well is then computed. Value of β versus s_m' are plotted on semilogarithmic paper, as shown in Figure 9. The best-fit straight line is constructed through the origin and these points, and its slope m_m ($\Delta s_m'$/cycle) is measured as 2.32; the line through the origin of steepest slope is also constructed. An average value of T is computed from $T = (2.3Q/2\pi m_m) = (2.3)(1.55)/(2\pi)(2.32) = 0.244$ ft^2/sec. The storage coefficient is then computed from $S = 4\pi T^2 m_x/Qz = 4\pi(0.244)^2(0.53\times60)/(1.55)(140) = 0.11$. For a minimum value of T, the slope of the enveloping line is to be used in computing for T and S (Fig. 9).

References

Jacob, C. E., *Engineering Hydraulics,* John Wiley & Sons, New York, ch. 5, 1950.

Kazmann, R. G., Notes on determining the effective distance to a line of recharge, *Trans. Am. Geophys. Union, 27,* 584–859, 1946.

Kazmann, R. G., The induced infiltration of river water to wells, *Trans. Am. Geophys. Union, 29,* 85–92, 1948.

Rorabaugh, M. I., Groundwater in northeastern Louisville and Kentucky with reference to induced infiltration, *U. S. Geol. Surv. Water Supply Pap. 1360-B,* 169 pp., 1956.

Stramel, G. J., C. W. Lane, and W. G. Hodson Geology and ground-water hydrology of the Ingalls area, Kansas, *State Geol. Surv. of Kansas Bull. 132,* 45–51, 1958.

Wisler, C. O., and E. F. Brater, *Hydrology,* John Wiley & Sons, New York, ch. 7, 1951.

Wenzel, L. R., Methods of determining permeability of water bearing materials, *U. S. Geol. Surv. Water-Supply Pap. 887,* 192 pp., 1942.

(Manuscript received July 17, 1959.)

Investigation of Water-Table Response to Tile Drains in Comparison with Theory

T. Talsma and Henry C. Haskew

C.S.I.R.O., Irrigation Research Station
and
Water Conservation and Irrigation Commission
Griffith, New South Wales, Australia

Abstract—An investigation of the performance of tile laterals, selected from farm drainage systems, is reported. Useful theories of water-table response to tile lines are briefly reviewed. The field procedure used for the investigation is described.

Average hydraulic conductivity was determined from in-place measurements by the auger-hole and piezometer methods. The position of the impermeable layer was determined from textural examination of the soil and from piezometer measurements of hydraulic conductivity. The performance of laterals was investigated by simultaneous measurements of water-table height along lines at right angles to the tile and rate of discharge from the tile line.

It is concluded that Hooghoudt's theory is adequately supported where flow boundaries can be sharply defined. The field data also support Kirkham's analysis, where the physical assumption underlying this analysis is reasonably met. Field data on the rate of lowering of the water table generally support Glover's analysis, although some caution appears to be necessary when using his analysis for design in cases where there is an impermeable layer at a small distance below the tiles.

Comparison of the data with theory also shows that average values obtained of factors used in design and those indicating performance are satisfactory, though field variability is high. The adaptability of field data to theoretical simplifications is emphasized.

Introduction

In the Murrumbidgee Irrigation Areas of New South Wales the auger-hole method described by *Maasland and Haskew* [1957] has been used, over the period from 1954 to 1957, to measure hydraulic conductivity on some 5000 acres of horticultural farm land. Tile lines have been installed during the same period on more than half of this area with spacings calculated from the formula of *Hooghoudt* [1940].

The performance of a number of tile lines has been studied in considerable detail for periods of more than a year. These tile lines were selected from farm tile-line systems designed on the basis of Hooghoudt's formula and from earlier tile-line systems designed on an empirical basis. The method of study is described. Data obtained have been used to check the physical response of the water table to tile systems and are compared with the theory of Hooghoudt as well as with the theories of *Kirkham* [1958] and Glover [*Dumm,* 1954].

Review of Theories

Maasland [1956] and *Van Schilfgaarde and others* [1956] fully review the available theories. Both conclude after considerable discussion that the steady-state solution obtained by *Hooghoudt* [1940] would give reliable results; under somewhat more restricted conditions Glover's nonsteady-state solution is also reliable [*Dumm,* 1954]. The assumptions of each, together with a recent analysis of *Kirkham* [1958], are presented briefly.

Hooghoudt's drain-spacing formula—In his analysis Hooghoudt assumes uniform downward percolation q from the soil surface to a static water table; that is, the flow rate and boundaries are independent of time. His formula is [*Hooghoudt,* 1940, p. 593]:

$$s^2 = 8(k/q)d(m - h) + 4(k/q)(m^2 - h^2) \quad (1)$$

where

k = hydraulic conductivity (ft/day)

q = rate of discharge per unit area of land surface (ft³/ft²/day)

d = $f(H, r, s)$, a function of the depth H of the impermeable layer below the level of the drains, the tile radius r, and the drain spacing s. The function takes account of convergence of flow near the tile (ft)

m = height of water above the tile lines at the point midway between two parallel lines (ft)

h = height of water immediately above a tile line (ft)

Published results [for example, *Kirkham and de Zeeuw*, 1952] indicate that for practical purposes h may be taken as zero. Formula (1) then reduces to

$$s^2 = 8(k/q)dm + 4(k/q)m^2 \qquad (2)$$

When, in addition, the impermeable layer is at tile level $H = 0$, formula (2) reduces to

$$s^2 = 4(k/q)m^2 \qquad (3)$$

When $H \gg m$, then also $d \gg m$ and the error introduced by neglecting the second term in (2) is small. The formula may then be reduced to

$$s^2 = 8(k/q)\ dm \qquad (4)$$

Kirkham's analysis—*Kirkham* [1958] gives an exact mathematical analysis of various flow configurations, based on the assumption $H \gg m$. This is tantamount to neglecting the loss of the hydraulic head of water above a plane going through the lowest point of the water table. One of his cases (water discharging into tile drains and no water over the tile lines—his case *d*) will be compared with some field data of th present investigation. The formula derived fo this flow problem is

$$m = \frac{sq}{k} F(H/s, 2r/s) \qquad (5$$

where

$$F(H/s, 2r/s) = \frac{1}{\pi}\left[\ln\frac{s}{\pi r} + \sum_{n=1}^{\infty}\frac{1}{n}\left(\cos\frac{2n\pi r}{s} - \cos n\pi\right)\left(\coth\frac{2n\pi H}{s} - 1\right)\right] \qquad (6$$

Kirkham's notation has been changed to cor form with the notation adopted in this pape (see formula (1) and Fig. 1).

Glover's formulas—Glover was primarily cor cerned with the disposal of excess irrigatio water during the interval between successiv irrigations. This is essentially a nonsteady-stat condition, and Glover, as reported by *Dumr* [1954], has proposed the following equatio (Fig. 2)

$$s^2 = \pi^2 \frac{kD_a t}{v \ln(4/\pi)(y_0/y_t)} \qquad ($$

where s and k are as previously defined, and

t = the time necessary to lower the wat table from an initial height y_0 to a fin. height y_t (days)

v = drainable or effective porosity per cer

D_a = average thickness of the aquifer (ft)

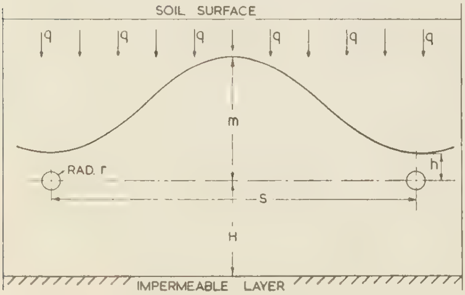


Fig. 1—Hooghoudt's flow geometry.

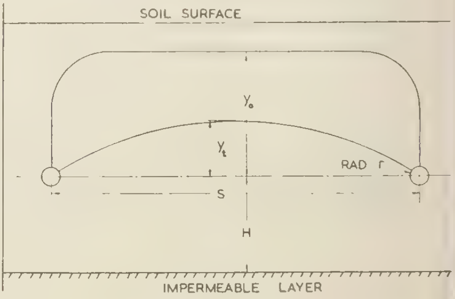


Fig. 2—Glover's flow geometry.

Dumm [1954] takes D_a to be a constant.

$$D_a = H + y_0/2 \qquad (8)$$

Formula (7) has been derived on the assumption of horizontal flow towards the drains. For large values of H flow convergence towards the drain becomes important; therefore *Maasland* [1956] has proposed the use of

$$D_a = d + y_0/2 \qquad (9)$$

where d is as in (1), to take account of this convergence.

To treat D_a as a constant further requires that y_0 must be small compared with H, which assumption is not justified if the impermeable layer approaches the level of the drains. For the case that $H = 0$ Glover therefore derived

$$s^2 = \frac{9ky_0t}{2v[(y_0/y_t)-1]} \qquad (10)$$

Dumm [1954] reported that (7) and (10) agree in practice to within 10 per cent and therefore (7) may be applied to cases where the impermeable layer is close to the level of the drains.

For $H = 0$, (7) and (10) are almost identical ($\pi^2 \simeq 9$ and $y_0/2 = D_a$) except for terms involving y_0/y_t. If we plot $(y_0/y_t) - 1$ against $\ln(4/\pi)$ (y_0/y_t) we find these terms to be equal only for $y_0/y_t = 1.85$. Agreement between (7) and (10) is closest near this value (that is, for moderate drawdowns). Dumm himself proposes to limit the use of (7) or (10) to moderate drawdowns. He reasons that the assumption, $y_0/2 = D_a =$ constant when $H = 0$ or is small, is not justified for large drawdowns. Even though both formulas will give nearly the same result when used for design in the restricted manner shown above, the discrepancy must be kept in mind when comparing them with field data.

Field Procedure

Farms which were already tile drained were selected for investigation. It was determined that the selected farms had the tile lines laid out in a rectangular pattern. The field data that were collected could then be better compared with theoretical results. Only farms with relatively low variation in hydraulic conductivity were included in the investigation. Average values of hydraulic conductivity differed from farm to farm, so the range normally found in routine investigations was adequately covered.

Investigation of each farm was in two parts which proceeded conjointly. The originally used values of factors determining depth and spacing (that is, the design factors k and H) were checked for accuracy by taking additional readings. Factors indicating the performance of the tile-line systems were observed over a lengthy period of time (6 to 18 months).

Design factors—Depth and spacing of most tile lines were originally determined by using Hooghoudt's formula in the manner described by *Maasland and Haskew* [1957]. Additional measurements of hydraulic conductivity were made by the auger-hole method or, where the soil above the impermeable layer was stratified, by the piezometer method. Hydraulic conductivity in the capillary fringe was neglected.

The impermeable layer, as defined by *Maasland* [1956, p. 5], was located by relating textural examination to hydraulic conductivity according to the method of *Talsma and Flint* [1958]. Where practicable, this method was supplemented by piezometer measurements.

Inside and outside diameters of tile used for laterals were 4 and 5½ inches, respectively. No effort was made to enclose the tiles in highly permeable material. A mean tile radius of 0.2 ft is therefore used in calculations with formulas (1) and (5). *Kirkham* [1958] has shown (see formulas (5) and (6)) that the radius r does not enter into the theoretical formulas in a sensitive way. Since the recently backfilled trench will be more permeable than the surrounding soil, it is reasonable to assume that there is no constriction of flow of water into the tiles.

Undisturbed core samples were collected from the farms studied, and drainable porosities were determined from these by applying 60 cm of suction on the tension plate. Auger-hole and piezometer measurements of hydraulic conductivity were made on the sites chosen for the core sampling. The 60-cm tension values, not being necessarily correct for all soils, were checked by readings of discharge volume and corresponding water-table fall. A drop, for example, of 1 ft per week in the water-table level with a discharge rate of 0.05 ft^3/ft^2/week indicates a drainable porosity of 5 per cent. Only periods with no appreciable evaporation or infiltration to affect the position of the water table were used for this comparison. Upward or downward seepage is negligible everywhere because of the presence in the deep subsoil of a layer of very low hydraulic conductivity. The difference between hydraulic head at top and bottom of this layer is also low. All water drained from the soil is thus assumed to be discharged from the tile lines.

Performance factors—Two or three lines of observation wells were placed at right angles across each selected lateral. Each well was a 7-ft length of 1¼-inch piping (22-gauge, hessian-wrapped, perforated) which was packed with gravel into a 2½-inch auger hole of about 6½-ft depth. The wells were placed one tree-row apart (approximately 22 ft). One well in each line was placed directly over or immediately beside the tile. The exact depth of tile line at this well was noted. The height of the top of each well was reduced to a standard level.

The depth from the top of the pipe to the water level was measured to 0.1 ft with an ordinary probe. The height of the water above any chosen reference level could then be calculated. During the present investigation the tile lines were seldom observed to run more than a quarter full. The plane at this level, rather than the center of the tiles, was thus taken as the reference level. In all cases averages of water-table height were obtained from several readings. Rainfall was measured by a gauge placed near each lateral.

Where possible, discharge rates were measured at individual laterals by timing a small quantity of flow into a graduated container. Elsewhere over-all discharge from an entire farm system was measured with a meter and a clockwork-driven recording chart. Values obtained were used in the comparison of field data with formulas (1), (5), (7), and (10) and also to check the standard value of the discharge coefficient, used for design, in (2). The standard value is 1/60 ft^3/ft^2/day when the water table is 1.5 ft (that is $m = 4.5$ ft) from the surface at the midpoints [*Maasland and Haskew,* 1957 Section VI. 1].

Analysis of field Data and Comparison with Theory

Some observations are first made on certain features of the design and performance factors. Complete data in regard to these factors are not given, but those presented will show the degree of natural variation that occurs and the manner of deriving average values for the vari-

Table 1—*Details of laterals investigated*

Farm No. and Lateral	Length, ft	Depth, ft	Spacing s, ft	H, ft	d, ft	k, ft/day	q* for $m = 2$ ft, mm/day
1298D	660	6.0	176	3.0	2.90	7.30	3.15
1235E	363	6.3	165	5.7	5.00	5.50	2.60
1303C	759	5.6	206	3.4	3.25	3.04	1.30
1185B	462	6.0	242	6.0	5.45	2.30	
1185C	462	6.0	154	6.0	5.10	2.30	
1185D	462	6.0	176	6.0	5.20	2.30	
1246B	726	6.1	137	Variable		1.66	1.10
853B	660	5.7	132	0.5-4.5	2.60	1.11	0.40
78B	1155	4.7	154	2-7	3.90	0.90	0.57
78C	1155	4.5	168	2-8	4.10	0.90	0.44
659E	578	6.4	132	2.6	2.50	0.55	0.39
76D	484	4.1	96	1.9	1.80	0.30	0.50
76J	484	4.1	96	1.1	1.00	0.54	0.88
2380B	429	5.8	110	0.5	0.50	0.58	0.29
2380D	429	5.9	110	0.3	0.30	0.32	0.12

*Average values from field data.

TABLE 2—*Variability of k values*

Farm No. and Lateral	Number of determinations (auger-hole method)	Average k (arithm. mean), ft/day	Standard deviation	Coeff. of variation, percentage
2380B	15	0.58	0.35	60
2380D	10	0.32	0.24	75
1303C	22	3.04	1.53	50
1298D	27	10.40	8.17	79

ous factors. The field data are then compared with results from theory.

Design factors—Details of laterals studied are summarized in Table 1. The symbols for the column headings are as in formula (1) and Figure 1. Results of individual auger-hole and piezometer measurements of hydraulic conductivity for two laterals (2380B and 76D) are given in Figure 3 in relation to the depth below soil surface at which the measurements were taken. Even in relatively homogeneous soil, there is often a significant relation between these two factors [*Talsma and Flint*, 1958].

From Figure 3 it is seen that the variation in k with depth for lateral 2380B was more or less at random to 6 ft below the surface, at which depth an abrupt decrease in k occurred. Some indication of a gradual decrease in k with depth can be seen for lateral 76D, though readings were taken at generally similar depths. The impermeable layer was approximately 6 ft below the surface for both laterals. Thus, the value of H and k for lateral 76D are 1.90 ft (that is, depth to impermeable layer, 6 ft, minus depth to reference level of tile line, 4.1 ft) and 0.30 ft/day.

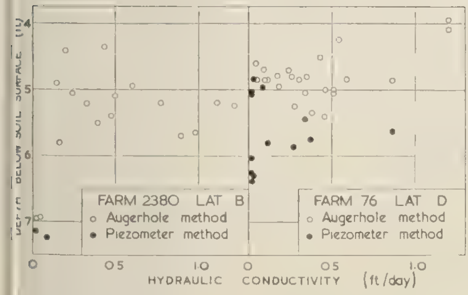


FIG. 3—k-values for laterals 2380B and 76D.

Irregular distribution of hydraulic conductivity in the vertical plane prevented reliable estimation of H, d, and k for some laterals (853B, 78B and C, and 1246B) of Table 1. Depth to the impermeable layer for these laterals varied considerably, with variation being as much as 0.5 to 8 ft below the tile level. A weighted average depth to the impermeable layer was calculated. Laterals 78B, 78C, and 853B were found to be partly installed in soil of lower-than-average hydraulic conductivity.

The average value of hydraulic conductivity above the impermeable layer was calculated for all laterals of Table 1 as the arithmetic mean of values remaining after any that were widely divergent were discarded [*Maasland and Haskew*, 1957, Sections VI. 2 and 3]. Table 2 gives some averages with their standard deviations; in calculating the averages of this table it was necessary to discard widely divergent k values only for lateral 1298D.

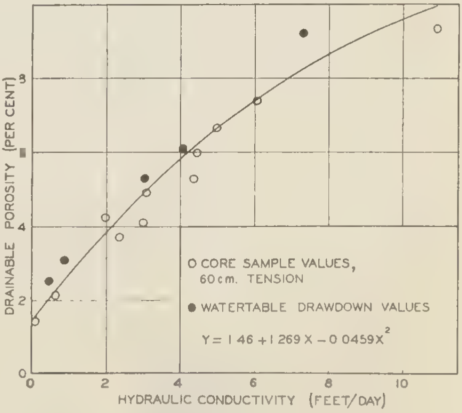


FIG. 4—Relationship between k and v.

Data could not be obtained in the above manner for lateral 1235E. Here there proved to be two distinct layers surrounding the lateral. From 3 to 9 ft the soil is a loam with average hydraulic conductivity of 3.10 ft/day; from 9 to 12 ft there is a sandy layer with average hydraulic conductivity of 10.40 ft/day. There are also sand lenses at 6 to 9 ft which appear to compensate for the decrease in k with depth of the loam. This part of the profile is probably anisotropic. The average k of lateral 1235E, used in Table 1, was calculated from these data, using the formula $kH = k_1H_1 + k_2H_2$. In this case $9k = 6 \times 3.10 + 3 \times 10.40$ and thus $k = 5.5$ ft/day.

The relationship between k and v for the core-sample and drain-discharge methods of estimation is shown in Figure 4. The core-sample method appears to give somewhat lower values of v. Statistically the relation is estimated differently by the two methods (test of displacement is significant at $P < 0.05$, but deviation from parallelism is not significant). Data from both methods have been combined in the one regression equation (Fig. 4). This equation should not be used outside the range of experimental data.

Performance factors—The observation-well readings of water-table level at points midway between the tile lines m were plotted against the corresponding discharge rate q for all laterals. Results for laterals 2380B and 1303C ar presented in Figures 5 and 6, respectivel Each point represents one reading of q and a average of 6 readings of m. Standard deviation of m were high.

Formula (2) may be written $q = \alpha m + \beta m$ where $\alpha = 8kd/s^2$ and $\beta = 4k/s^2$. Regressio equations of this form have been fitted to th data of Figures 5 and 6, with the symbols and β' being used to indicate regression value The values of α, α' and β, β', are included i the figure legends and are respectively of th same order of magnitude. The ratio α'/β' in creases as the impermeable layer recedes belo the tile lines (this is not surprising since th ratio $\alpha/\beta = 2d$, which quantity is primaril determined by H; see Table 1). The value c the quotient α'/β' might therefore be used t estimate the approximate position of the im permeable layer. Some caution must be exer cised however; the above laterals were no selected arbitrarily, but as cases where ther is little decrease of k with depth. Where k doe decrease with depth the quadratic term in th regression equation will become relatively large

Extrapolation of the curve of Figure 6 show that discharge from lateral 1303C is close to th standard value of 1/60 ft^3/ft^2/day (that is 5 mm/day) at $m = 4.5$ ft. Actual and recom mended spacings for this lateral are within few feet of each other. Figure 5 shows the actua

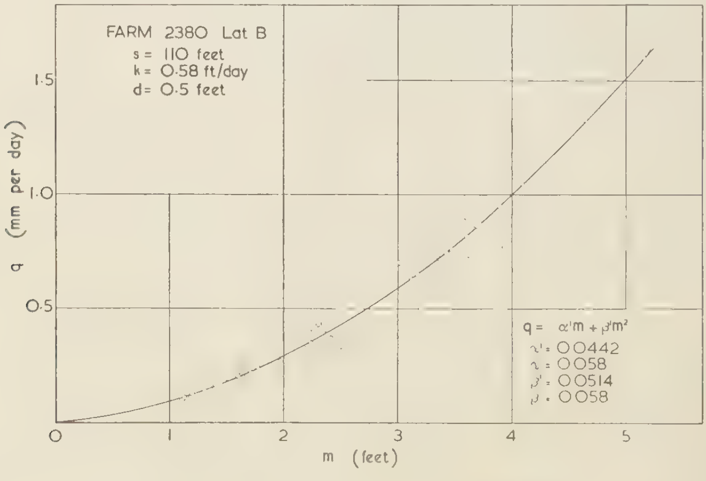


FIG. 5—Relationship of q and m for lateral 2380B.

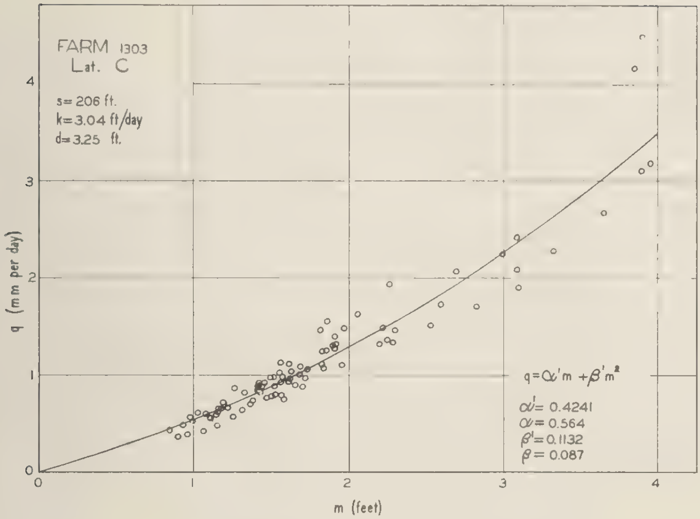


Fig. 6—Relationship of q and m for lateral 1303C.

discharge of lateral 2380B to be 1.25 mm/day at m = 4.5 ft. Actual spacing for this lateral is 110 ft, which is almost double the recommended spacing of 56 ft, as calculated with (2) with $q = 5$ mm/day and $m = 4.5$ ft. As the design factors k and d remain the same for this lateral for any drain spacing s, the product s^2q is a constant for both spacings, and thus the actual discharge for the recommended spacing of 56 ft may be calculated. We find $q = 1.25 \times 110^2/56^2 = 4.82$ mm/day, which is very close to the standard value of 5 mm/day.

Some reasons can be given to explain the scatter of points in Figures 5 and 6, apart from the high standard deviations of m and errors of observation. Abnormally high discharges were found to occur immediately after irrigation or heavy rainfall when the water table was rising rapidly; for example, the two highest of lateral 1303C (Fig. 6). Some low discharges occurred with the sharp drop of the water table soon after irrigation or rainfall. This was especially so when the water table was fairly high. A check on the observation points close to the regression line showed nearly all these to be taken at times when the water table was approximately steady, slowly rising, or slowly dropping. Only for such dates have observations been used in the comparison of field data with Hooghoudt's formula, which, it is remembered, has been derived for equilibrium conditions.

The last column of Table 1 contains values of q corresponding to $m = 2$ ft for all laterals except 1185B, C, and D, for which discharge records from individual laterals were not possible. Values of h, corresponding to $m = 2$ ft are not given; these varied from 0 to 0.4 ft for the various laterals. Discharge values in Table 1 were obtained from the regression equations, examples of which are included in Figures 5 and 6. The relationship between the values of q and k was found to be essentially linear ($r = 0.97$, $P < 0.001$), in agreement with (1), (5) and (7). Deviations from linearity that have occurred in the above data are explained as follows. (a) The data have not been reduced to a common flow geometry (say, $H = 3$ ft and $s = 150$ ft) and it is not possible to do this. (b) When $m = 2$ ft the water table above the shallow laterals 76D and J is only 2 ft below the surface. Some flow then occurs through the more permeable top soil, thus rendering inaccurate the value of k.

The shape of the water table across two laterals (2380B and 1235E) is shown in Figure 7. Since spacings differ the coordinates are shown as $2x/s$ and $2y/s$. The origin of the coordinates is the point of intersection of the

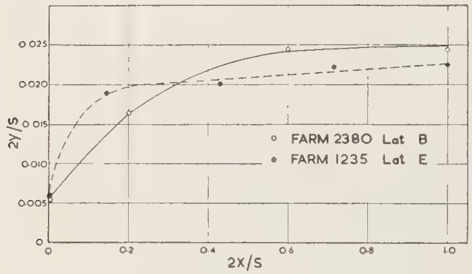


FIG. 7—Watertable shape laterals 2380B and 1234E.

vertical through the tile center and the reference plane. Readings used for Figure 7 are for similar water-table heights but not for the same dates. It is pointed out, however, that the shape of the water table for any lateral appeared independent of water-table height and thus also of time. The rather flat shape of the water table across lateral 1235E may be taken as an indication of the existence of anisotropy [se under 'Design factors,' *Maasland*, 1957].

Small values of the height of water immediately over the tile lines *h* were measured in mos cases (Fig. 1 and Table 3). This may be due t the tile not being embedded in highly permeabl material. Very small values of *h* were difficul to determine accurately. Points of inflection i the shape of the water table, found in theoretical solutions [for example, *Van Deemter*, 1950 *Childs*, 1947] were not detected.

Comparison with theory, (i) *Hooghoudt'* *drain-spacing formula*—Data for all laterals except 1246B for which *H* could not be reliabl estimated, will be compared with Hooghoudt' formula. Values of *q*, *m*, and *h* for the selecte dates (Table 3) were substituted in (1) togethe with *d* and *k* values from Table 1. The m^2 an h^2 values in the formula are the means of sum of squares of the replicates (4 or 6 values of *m* and 2 or 3 values of *h*).

Table 3 is divided into sections A, B, and C

TABLE 3—*Comparison of field data with Hooghoudt's formula*

Section	Farm No. and Lateral	Date	*q*, mm/day	*m*, ft	*h*, ft	Spacing *s*, ft Actual	Spacing *s*, ft Calculated	Mean,* ft	Deviation percentag
A	1303C	7/20/56	2.50	3.02	0.52	206	194	205 (6)	−0.5
		5/14/56	1.12	1.84	0.17		218		
	2380B	6/20/56	0.65	3.17	0.67	110	114	117 (7)	+6.4
		4/26/56	0.24	1.76	0.42		117		
	2380D	10/30/56	0.45	3.48	0.57	110	118	126 (4)	+14.5
		4/5/57	0.16	2.12	0.22		132		
	659E	9/4/56	0.28	1.61	0.49	132	137	137 (1)	+3.8
	76D	2/12/57	0.252	1.43	0	96	104	122 (3)	+27.1
		1/8/57	0.113	1.10	0		129		
	76J	1/17/57	0.238	1.12	0	96	98.4	93.5 (3)	−2.6
	1185C	8/13/56	1.10	1.00	0	154	169	169 (1)	+9.7
	1185D	8/13/56	1.10	1.20	0	176	188	188 (1)	+6.8
	1185B	8/13/56	1.10	1.84	0	242	244.5	244.5 (1)	+1.0
B	78B	8/10/56	0.96	2.58	0.45	154	162	178 (5)	+15.6
		10/19/56	0.48	1.65	0.17		184.5		
	78C	6/18/56	0.40	1.81	0.20	168	212.5	222 (3)	+28.7
	853B	3/8/56	0.76	2.36	0.70	132	163	183 (3)	+38.5
C	1235E	10/22/56	2.85	2.00	0.55	165	207		+25.4
	1235E†	10/22/56	2.85	2.00	0.55	165	155		−6.1

*Mean determined from number of calculations indicated in brackets.
†Calculated with $d = 5$ and $k = 3.10$.

he laterals of section A have sharply defined ow boundaries and average hydraulic conictivity. In section B the flow boundaries are ss distinct and the distribution of k values in e vertical plane is irregular. Lateral 1235E ection C) has already been discussed.

The data of Table 3, section A, are substanally in accord with Hooghoudt's claim of : 10 per cent accuracy for results with his rmula. Only for lateral 76D is there considerole difference between actual and calculated oacings. The results of Figure 3, however, dicate the presence of a layer of low permeality slightly below tile level for this lateral. Its resence would reduce the flow towards the le line somewhat and could be accounted for y selecting a lower value for k. No logical basis r such selection is known to the authors. The 10 mm/day discharge for laterals 1185B, C, id D could not be measured directly but is elieved to be a reliable estimate from the availole figures.

It will be noted that calculated spacings tend increase when m decreases; that is, when ie water table drops. This may be partly cplained by the decrease of hydraulic conictivity with depth [*Talsma and Flint,* 1958]. he general trend is for the calculated spacings be higher.

Calculated spacings in Table 3, section B, are o large, especially for lateral 853B. This is not irprising, since the vertical distribution of k such that most low values occur at the tile vel. The average value of k used for lateral 53B is somewhat too large, since not enough easurements of hydraulic conductivity in the ss permeable subsoil were available. While the ita on these laterals should not be taken as idence to prove that Hooghoudt's analysis is error, it is pointed out that his formula, or r that matter any of the other available formulas, should be used with caution in such cases.

Table 3, section C, shows that, if the average k as determined previously (Table 1) is used, the influence of the more permeable layer (which occurs at 2½ ft below the tile level) is very much overestimated. If a spacing is calculated with the k value of 3.10 ft/day in the upper layer in which the tile line is placed, the error is reduced considerably. This is in agreement with *Kirkham* [1951] who, for similar cases with a horizontal water table, concluded that the hydraulic conductivity of the soil in which the tile was placed almost entirely dominated the flow. The same conclusion was reached by *Van Deemter* [1950].

Variation between individual auger-hole values of k for lateral 1298D was very high (Table 2). The values are distributed at random both horizontally and vertically. Mode of distribution of values between the extremes (1.5 to 32.0 ft/day on the one soil type) made it difficult to select an average value by discarding those that were widely divergent. The value of 7.3 ft/day (Table 1) thus derived deviates considerably from the arithmetic mean of 10.4 ft/day (Table 2).

Hooghoudt's formula may be used here to calculate an average hydraulic conductivity, since the occurrence of the impermeable layer is quite sharply defined and sufficient values of q, m, and h are recorded under approximately steady-state conditions. Some of these values are given in Table 4. The mean value (6.9 ft/day) from Table 4 is considerably lower than the arithmetic mean (10.4 ft/day) of Table 2. The true mean value will be between 5.8 and 8.3 ft/day if Hooghoudt's 20 per cent limits of accuracy for determining k with (1) are accepted.

Thus it appears that when k values are dis-

TABLE 4—*Calculation of average k for lateral 1298D with formula (1)*

Date	q, mm/day	m, ft	h, ft	k, ft/day	Mean* k, ft/day
9/4/56	1.18	0.82	0.22	7.3	6.9
12/7/56	2.13	1.50	0.43	6.6	
7/24/56	2.84	1.77	0.43	6.8	

*Mean is average of 7 calculations.

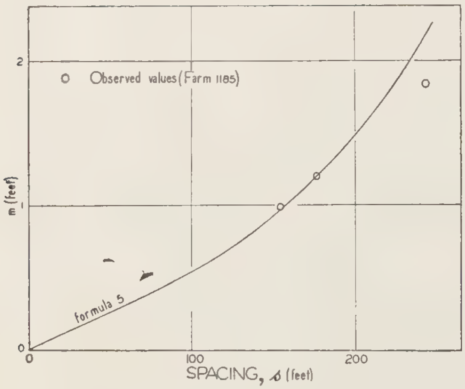


Fig. 8—Comparison of field data with formula (5).

tributed at random the arithmetic mean may lead to average values that are too high. Spacings calculated from the discharge and such average values of k should be higher than actual spacings. Table 3, section A, shows that this trend is apparent. When k decreases with de (for example, 2380D and 76J) the lower va near the tile would also tend to result in hig calculated spacings. Neglect of the contribu of the capillary fringe to flow [see, for exam *Van Schilfgaarde and others,* 1956, p. 671] ı have partly compensated for the above err

(ii) *Kirkham's analysis*—The data on three laterals 1185 in Table 3, section A, ı also be compared with Kirkham's theory, s the asumption underlying his analysis (tha $H \gg m$) is more nearly met in this case t in all others. The result of this compariso shown in Figure 8. Values of $F(H/s, 2r/s)$, ι in computing the theoretical curve, have b obtained by graphical interpolation of va given by *Kirkham* [1958, Table 1]. It is s that the agreement is good. The discrepa between theory and field data for the lar spacing is quite understandable because I H is only about three times the value of m

Kirkham's analysis of his case d provide

Table 5—*Comparison of field data with formulas (7) and (10)*

Farm No. and Lateral	Period	t, days	H, ft	D_a, ft	y_0, ft	y_t, ft	v	Spacing, ft Actual	Calculated, formula (7)	Calculated, formula (10)	Deviation, percentage Formula (7)	Forn (1(
1185C	8/9–16/56	7.22	6.0	5.75	1.30	0.77	0.042	154	171		+11.0	
1185D	8/9–16/56	7.22	6.0	5.97	1.53	0.97	0.042	176	184		+4.5	
1185B	8/9–16/56	7.22	6.0	6.54	2.18	1.58	0.042	242	213		−12.0	
1303C	8/7–16/56	8.84	3.4	4.98	3.09	1.96	0.049	206	197		−4.4	
1298D	8/9–16/56	6.96	3.0	4.13	1.96	1.23	0.083	176	188		+6.8	
76D	11/21–27/56	6.00	1.9	2.36	1.11	0.99	0.019	96	79	65	−17.7	+3
	7/17–19/57	1.88	1.9	2.95	2.31	2.15	0.019		52	64	N.D.	N
	7/17–22/57	4.88	1.9	2.95	2.31	1.81	0.019		68	54	N.D.	N
76J	11/21–27/56	6.00	1.1	1.41	0.82	0.69	0.021	96	72	55	−25.0	−4
2380B	3/14–19/56	5.00	0.5	2.38	3.29	2.59	0.022	110	80	85	−6.4	+
	8/8–17/56	8.73	0.5	3.10	4.37	3.59	0.022		126	144		
2380D	9/11–14/56	3.00	0.3	1.92	2.83	2.71	0.019	110	58	121	−47.3	+1

erification of (4), and values of m obtained ith each should agree closely. For $s = 154$, 76, and 242 ft, the values of m are 0.94, 1.20, 16 and 0.91, 1.17, 2.10, respectively. Slightly gher values are obtained with Kirkham's rmula. Therefore, spacings calculated with ooghoudt's formula (4) would be slightly rger.

(iii) *Glover's formulas*—Comparison of field ata with (7) and (10) is contained in Table Values of v are taken from the regression uation of Figure 4. Only laterals (from Table section A, plus 1298D) with well-defined flow ometry have been used. Laterals with largest alues of H appear first. D_a was given by (9); here water stood over the tile, $D_a = d +$ $y_0 + h_0)/2$ is perhaps a better approximation nd was used in these cases. Similar precauons, as taken for the data in Figure 4, on the hoice of periods for this comparison were necesry. Such periods could not be obtained for teral 2380B, and during the first period shown ere was very likely some evaporation, whereas uring the second period there was some infilation of temporarily stagnating water on the early impervious B-horizon of this particular il. The percentage differences calculated for is lateral are for the mean of the two callated spacings.

For larger values of H agreement of field data th formula (7) is good. The use of $d =$ $H, r, s)$ instead of H gave better agreement here d and H differed considerably (1185 C nd D). Disagreement of field data with (7) in ses where the impermeable layer approaches e tile level (laterals 76D and J, 2380B and , where $H \approx y_0$) indicates that here (7) does t accurately describe the lowering of the ter table. The value of y_0/y_t was much less an 1.85 in these cases.

The field data on these laterals have also en compared with Glover's second equation 0) and it was assumed for this purpose that $= 0$ in each case. Table 5 shows that there good agreement in the case of laterals 2380B d D, where the impermeable layer is close to e tile. The lack of agreement for laterals 76D d J, where $H \approx y_0$, is apparently due to an derestimation of the average thickness D_a of e aquifer.

It may be concluded that the field data generally support Glover's analysis, although some caution appears to be necessary when using his analysis for design of drainage systems in cases where there is an impermeable layer at a small distance below the tiles. Formula (7) might here be used, provided that only moderate drawdown is required. Where the impermeable layer is very close to, or at the level of, the tiles, use should be made of (10). The main objection of *Van Schilfgaarde and others* [1956], the neglect of convergence, appears to be adequately overcome if H is replaced by $d = f(H,r,s)$.

Discussion

The theoretical assumption of a soil of uniform hydraulic conductivity to a certain depth, at which an impermeable layer occurs, is not easily realized in the field (Fig. 3, Table 2). Reduction of field data to this simplified picture is often feasible and gives satisfactory results in the design of tile drainage installations.

Evaluation of an installed system is also feasible, though a large number of replicates appear necessary to measure average height of water table and discharge. The former can be achieved for a single lateral by using several measuring points; the latter can only be obtained in time, by measurements at similar water-table heights. *Kirkham and de Zeeuw* [1952] found a similar variability in a replicated drainage experiment on what is described as a very uniform site. These authors also concluded that the averaging of a large amount of data should give reliable results.

Data and results in Tables 1 and 3 indicate that approximately 30 per cent of the laterals investigated were installed in soil that was too heterogeneous for close agreement of tile performance with theory to be obtained. This situation could be improved upon in some cases by setting out the tile lines in a more complicated design, although very complicated designs should be avoided for practical reasons. The authors' experience is that soil heterogeneity requires caution in the use of drain-spacing formulas in about 25 per cent of the farms recommended for tile drainage in the Murrumbidgee Irrigation Areas.

In general the field data presented agreed favorably with theory wherever the assumptions underlying the theoretical solutions were reasonably fulfilled.

References

Childs, E. C., The watertable, equipotentials and streamlines in drained land: V. The moving watertable, *Soil Sci., 63,* 361–376, 1947.

Dumm, L. D., New formula for determining depth and spacing of subsurface drains in irrigated lands, *Agr. Eng., 35,* 726–730, 1954.

Hooghoudt, S. B., Bijdragen tot de kennis van eenige natuurkundige grootheden van den grond, 7, Algemeene beschouwing van het probleem van de detail ontwatering en de infiltratie door middel van parallel loopende drains, greppels, slooten en kanalen, *Verslag. Landbouwk. Onderzoek., 46,* 515–707, 1940.

Kirkham, Don, Seepage into drain tubes in stratified soil, *Trans. Am. Geophys. Union, 32,* 422–442, 1951.

Kirkham, Don, Seepage of steady rainfall through soil into drains, *Trans. Am. Geophys. Union, 39,* 892–908, 1958.

Kirkham, Don, and J. W. de Zeeuw, Field measurements for tests of soil drainage theory, *Soil Sci. Soc. Am., Proc., 16,* 286-293, 1952.

Maasland, M., The relationship between permeability and the discharge, depth and spacing of tile drains, *Water Conserv. Irrig. Comm. N. Wales Bull. Groundwater and Drainage Ser. n 1,* 35 pp., 1956.

Maasland, M., Soil anisotropy and land drainag in *Drainage of Agricultural Lands,* J. N. Luthi ed., ch. 2, V, American Society of Agronom Madison, Wisconsin, 1957.

Maasland, M., and H. C. Haskew, The augerho method of measuring the hydraulic conductivi of soil and its application to tile drainage pro lems, *Proc. 3rd. Intern. Cong. Irrig. Drainag* 8.69–8.114, 1957.

Talsma, T., and S. E. Flint, Some factors d termining the hydraulic conductivity of su soils with special reference to tile draina problems, *Soil Sci., 85,* 198–206, 1958.

Van Deemter, J. J., Bijdragen tot de kennis va enige natuurkundige grootheden van de gron 11, Theoretische en numerieke behandeling va ontwatering—en infiltratie-stromingsproblem *Verslag. Landbouwk, Onderzoek., 56,* no. 7, 195

Van Schilfgaarde, Jan, Don Kirkham, and R. Frevert, Physical and mathematical theories tile and ditch drainage and their usefulness design, *Iowa Agr. Exp. Sta. Res. Bull. 436,* p 667–705, 1956.

(Manuscript received February 27, 1958; revis May 4, 1959.)

Direction of Polarization Determined from Magnetic Anomalies[1]

DONALD H. HALL

Geophysics Laboratory, University of British Columbia
Vancouver 8, B.C., Canada[2]

Abstract—In traditional methods of interpretation of the results of magnetic surveys the effects due to permanent magnetization are neglected. Recent geomagnetic research on the remanent magnetization of rocks has shown this neglect to be unjustified. Moreover, techniques now being employed provide better measurements of magnetic-field variations than have ordinarily been available in the past. With most of the newer methods components of the field are measured in directions other than those generally treated in the earlier methods of interpretation. In order to take advantage of the new developments, equations for the magnetic field over a point dipole, a horizontal line of dipoles, a thin, dipping sheet, a thick, dipping sheet, and a sloping step are derived in the case where both the directions of measurement and polarization are arbitrary.

It is found that these directions combine with other properties of the bodies to form parameters from which various features of the magnetic anomalies over the bodies can be determined. In terms of these combined parameters it is possible to obtain the higher derivatives of the expressions for the fields over these bodies and to develop methods of determining the unknown parameters of the bodies, including the direction of polarization, when magnetic profiles over them are given. Further, it is shown that the fields over four of the bodies treated can be obtained by successive differentiation of a single function. This fact is used in drawing charts for computing values of the fields and their derivatives at points along profiles over any of these bodies. Graphs are given, showing the position of special points such as peaks and inflection points on the profiles for any direction of polarization and measurement.

Methods of calculation of the unknown parameters of the line of dipoles and the thin, dipping sheet when their anomalies are given, are outlined, followed by examples of calculation in the case of theoretically calculated sample profiles.

Introduction—Many magnetic anomalies can be interpreted in terms of depth, size, and other quantitative features of magnetized formations in the subsurface, and a great many methods of carrying out the interpretation have been devised. It was formerly believed that for the most part the magnetization arises from induction in the earth's field. If this is the case, the magnetization can always be taken to lie along this known direction, and the susceptibility of the formations can be calculated without difficulty. This is done in many of the methods of magnetic interpretation [see, for example, *Heiland* 1946, pp. 389–400]. The effect of any component of permanent magnetization along this direction would be included in the value obtained, if cases were encountered where this type of magnetization occurs to a significant degree [*Heiland*, 1946, pp. 401–402]. Since the direction of the magnetization of a body, as well as its intensity, influences the size and shape of the associated anomaly, any appreciable amount of permanent magnetization in directions other than along the earth's field would give rise to anomalies requiring a more extended interpretation than is provided for in the simpler methods mentioned above. Furthermore, it is becoming increasingly clear that, contrary to the earlier belief, permanent magnetization is widespread in the rocks of the earth's crust.

This realization of the importance of permanent magnetization is one of the results of laboratory work on the magnetic properties of rocks. It has also been established that in many cases this type of magnetization is of the same order of magnitude or even many times greater

[1] Based on a thesis submitted as partial requirement for the degree of Doctor of Philosophy at the University of British Columbia.

[2] Present address: Department of Geology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

than that of the induced magnetization. Directions other than that of the earth's field are common, and these are found to be valuable indicators of the geological history of the rocks. The significance of this new understanding of the importance of permanent magnetization of rocks is well summarized by *Blackett* [1956]. Laboratory measurements of the direction of magnetization have been made on large numbers of specimens in order to draw conclusions about geological conditions and history, and any method which makes it possible to estimate this quantity from magnetic anomalies would supplement these as a valuable additional source of geological information.

It is our purpose in the present paper to outline a method of interpretation in which the direction of polarization can be calculated as an unknown from magnetic anomalies. Since it is desirable to utilize the results of modern aeromagnetic surveys, certain additional requirements of the instruments used in these surveys must be taken into account. In the older instruments, such as the magnetic balance, only the vertical and the horizontal components of the field were measured, and the effect of direction of measurement did not have to be considered in a general way. The airborne instruments used in modern surveys, however, may be oriented so as to measure the component of the field in any desired direction. Since the particular direction of measurement, like the direction of polarization, influences the size and shape of anomalies, any method of interpretation to be used in conjunction with modern airborne instruments should allow for a completely arbitrary direction of measurement. This is done in the method of interpretation presented here. It is found that treating both direction of measurement and direction of polarization as completely arbitrary leads to certain combined parameters and to methods of solution of wider applicability but of no greater complexity than those developed formerly for various particular cases of the directions.

Among the first treatments of direction of polarization as an unknown to be determined were those of *Mikov* [1953] and *Voskoboynikov* [1955], who considered the vertical and horizontal components of the field over horizontal, polarized cylinders. *Sutton and Mumme* [1957] considered the effect of arbitrary direction of polarization on the aeromagnetic anomalies over a single dipole and a line of dipoles. In the present paper the directions are taken to be entirely arbitrary, and direction of polarization is an unknown to be determined.

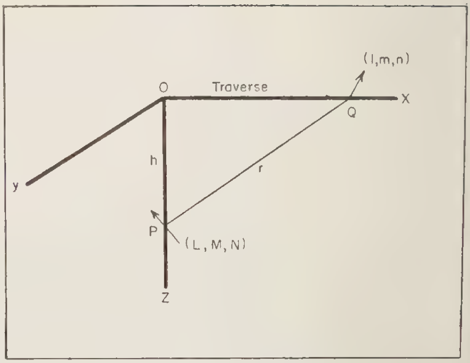


FIG. 1—Traverse over a single dipole.

The magnetic field over a point dipole (*Fig. 1*)—Let a magnetometer Q at the point (x, y, z) measure $\mathbf{F}$, the component in the direction (l, m, n) of the field due to a dipole P of moment μ, polarized in the direction (L, M, N) and located at the point (a, b, c). Then $\mathbf{F}$ is given by [*Jeans*, 1948, p. 372]

$$\mathbf{F} = \mu \frac{\partial^2}{\partial s\, \partial t}\left(\frac{1}{r}\right) \tag{1}$$

where $\partial/\partial t$ is differentiation in the direction (l, m, n), that is, in the direction of the component to be measured, $\partial/\partial s$ is differentiation in the direction (L, M, N), that is, in the direction of the north pole of the dipole, and

$$r^2 = (x - a)^2 + (y - b)^2 + (z - c)^2 \tag{2}$$

If f and g are any two functions of x, y, and z,

$$\partial f/\partial s = L\, \partial f/\partial x + M\, \partial f/\partial y + N\, \partial f/\partial z \tag{3}$$

and

$$\partial g/\partial t = l\, \partial g/\partial x + m\, \partial g/\partial y + n\, \partial g/\partial z \tag{4}$$

These expressions can be used to expand (1) in the form

$$\mathbf{F} = \frac{\mu}{r^5}\,[\alpha_{11}(x - a)^2 + \alpha_{22}(y - b)^2 + \alpha_{33}(z - c)^2 + \alpha_{12}(x - a)(y - b) + \alpha_{13}(x - a)(z - c) + \alpha_{23}(y - b)(z - c)] \qquad (5)$$

where

$$\begin{aligned} \alpha_{11} &= 2Ll - Mm - Nn & \alpha_{12} &= 3(Ml + Lm) \\ \alpha_{22} &= 2Mm - Nn - lL & \alpha_{13} &= 3(Nl + Ln) \\ \alpha_{33} &= 2Nn - lL - Mm & \alpha_{23} &= 3(Nm + Mn) \end{aligned}$$

Observations in a horizontal plane above a single dipole—Taking the z-axis downward, and the dipole at $(0, 0, h)$ we have the case of observation in the x-y plane, which is at a distance h above the dipole.

Thus in (5),

$$z = a = b = 0, \quad \text{and} \quad c = h \qquad (6)$$

and hence

$$\mathbf{F} = \frac{\mu}{(x^2 + y^2 + h^2)^{5/2}}\,(\alpha_{11}x^2 + \alpha_{22}y^2 + \alpha_{33}h^2 + \alpha_{12}xy - \alpha_{13}xh - \alpha_{23}yh) \qquad (7)$$

Thus the directions of polarization and measurement, as well as depth, are among the unknown parameters of a body which determine the shape of the corresponding anomaly. The various charts given by *Vacquier* [1951] illustrate this with prismatic bodies for the special case $(L, M, N) = (l, m, n)$.

If the axes are taken so that the y-axis is along the magnetic meridian, equation (7) is equivalent to equation (6) of *Sutton and Mumme* [1957]. In this form it is suitable for calculating the actual contours over a single dipole or, as is done by the above-mentioned authors, for studying the profiles in special directions, such as along or perpendicular to the magnetic meridian.

However, if the equations for the dipole are to be integrated to obtain profiles over more complex bodies, the more general equation (7) is preferable.

Observations along a traverse—If Q (Fig. 1) is confined to the x-axis, then $y = 0$. A further simplification results, particularly in performing the actual computation, from writing distances in units of h and going over to what we shall call the 'reduced form' of the expression for **F**. Applying these to (7) and letting $\xi = x/h$, we obtain

$$\mathbf{F} = \frac{\mu\alpha_{11}}{h^3(\xi^2 + 1)^{5/2}}\,(\xi^2 - C_1\xi + C_2), \qquad C_1 = \alpha_{13}/\alpha_{11}, \qquad C_2 = \alpha_{33}/\alpha_{11} \qquad (8)$$

Differentiating, we have

$$\frac{\partial\mathbf{F}}{\partial\xi} = -\frac{\mu\alpha_{11}}{h^3(\xi^2 + 1)^{7/2}}\,[3\xi^3 - 4C_1\xi^2 + (5C_2 - 2)\xi + C_1] \qquad (9)$$

and a second differentiation gives

$$\frac{\partial^2\mathbf{F}}{\partial\xi^2} = \frac{\mu\alpha_{11}}{h^3(\xi^2 + 1)^{9/2}}\,[12\xi^4 - 20C_1\xi^3 + 3(10C_2 - 7)\xi^2 + 15C_1\xi - (5C_2 - 2)] \qquad (10)$$

The shape of the profile, then, is determined by three quantities: h, C_1, and C_2. Maxima and minima occur at values of ξ for which $\partial\mathbf{F}/\partial\xi = 0$, and inflection points where $\partial^2\mathbf{F}/\partial\xi^2 = 0$.

If the position of the dipole in the horizontal plane can be determined by other methods, for instance, from a gravity survey or from geological indications, then the origin of the coordinate system in (8) to (10) may be fixed. If we use the positions of peaks and inflection points along one, or a number of profiles through the origin, we can determine values of L, M, N, and h without difficulty.

A single dipole approximates a compact, uniformly magnetized body at depth, with dimensions roughly the same in all directions. Such bodies are encountered especially in mining exploration [see, for example, *Yüngül*, 1956], and for this reason the equations developed for the dipole may on occasion be required for the solution of practical problems.

They are intended here, however, as the starting point for developing equations for the profiles over more complex bodies.

The horizontal line of dipoles (*Fig. 2*)—Let the magnetic moment have a constant line density

μ_L per unit length, and consider a straight, horizontal traverse perpendicular to the line and passing over it at a height h. If we take axes with z vertically downward, and the traverse along $0x$, then any element, length dt, of the line at $b = t$ can be considered a dipole with magnetic moment $\mu_L\, dt$, and it produces a field $d\mathbf{F}$ at the magnetometer, which is given by (5) in which $y = z = a = 0$, $b = t$, $c = h$, and $r^2 = x^2 + h^2 + t^2$.

Then

$$dF = \frac{\mu_L}{r^5}(\alpha_{11}x^2 + \alpha_{22}t^2 + \alpha_{33}h^2 - \alpha_{12}xt - \alpha_{13}xh + \alpha_{23}th)\, dt \qquad (11)$$

If $\mathbf{F}_{t_1}$ is the force due to a line terminating at $y = \pm\, t_1$

$$\mathbf{F}_{t_1} = \int_{-t_1}^{t_1} d\mathbf{F} \qquad (12)$$

Substituting from (11) into (12) and carrying out the integration, we find

$$\mathbf{F}_{t_1} = \frac{2t_1^{\,3}\mu_L}{(x^2 + h^2)^2(t_1^{\,2} + x^2 + h^2)^{3/2}} \cdot \left\{ \begin{aligned} &\frac{\alpha_{11}}{t_1^{\,2}}\, x^4 + \frac{3}{2}\frac{C_5}{t_1^{\,2}}\, hx^3 \\ &+ \left[C_4 + \frac{1}{t_1^{\,2}}(\alpha_{11} + \alpha_{33})\right]h^2x^2 \\ &+ \left(1 + \frac{3}{2t_1^{\,2}}\right) C_5h^3x \\ &+ \left(C_6 + \frac{\alpha_{33}}{t_1^{\,2}}\right)h^4 \end{aligned} \right\} \qquad (13)$$

where $C_4 = -C_6 = (lL - nN)$ and $C_5 = -2(lN + Ln)$

If $t_1 \to \infty$, we have the case of an infinite line of dipoles, and $\mathbf{F}_{t_1} \to \mathbf{F}$ which is given by

$$\mathbf{F} = 2\mu_L\left[\frac{C_4x^2 + C_5hx - C_4h^2}{(x^2 + h^2)^2}\right] \qquad (14)$$

We may write this equation in reduced form, dividing numerator and denominator by h^4 and factoring out C_4, in which case we have

$$\mathbf{F} = \frac{2\mu_L C_4}{h^2}\left[\frac{\xi^2 + \Lambda_1\xi - 1}{(\xi^2 + 1)^2}\right] \qquad (15)$$

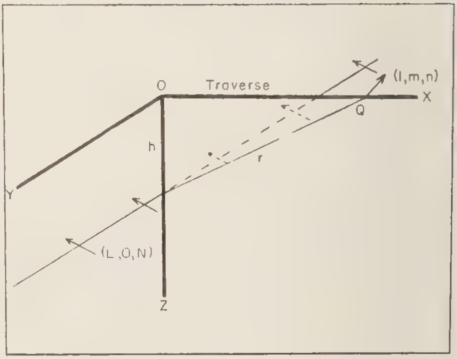


FIG. 2—Traverse over an infinite, horizontal line of dipoles.

where $\Lambda_1 = C_5/C_4 = 2(lN + nL)/(nN - lL)$. This may be called the 'polarization function.'

The magnetic field over bodies derived from the infinite line of dipoles—The equations for the magnetic field and derivatives over two elementary bodies—the single dipole and the line of dipoles—have been generalized to include cases where both the directions of polarization and measurement are arbitrary. Within reasonable limits of accuracy, such elementary forms do have a wide application in practical interpretation [*Nettleton*, 1942, p. 293], but if the methods are to be used to study the direction of polarization over a wide variety of geological situations, or if generalizations are to be made, a wider range of bodies must be treated. For this purpose, expressions for the field over a number of pris-

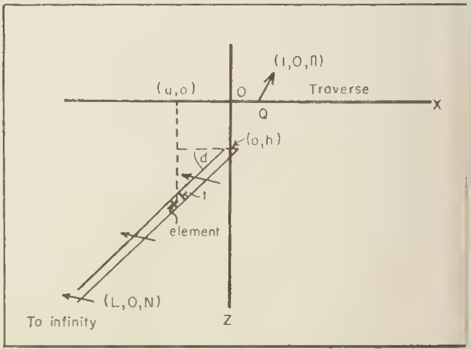


FIG. 3—Cross section of a thin, dipping, semi-infinite, polarized sheet, with element of integration.

matic, infinitely long bodies, assumed to have uniform polarization, will be derived from the results for the infinite line of dipoles.

The thin, dipping, semi-infinite polarized sheet—In Figure 3, a cross section of the body and of an element, chosen so as to approximate a line of dipoles, are shown. If the sheet has a uniform intensity of magnetization I, then referring to the figure, we see that its magnetic moment per unit length is It (sec d)du, while the depth of the line is $h + u \tan d$. Substituting into (14) we find

$$d\mathbf{F} = 2It\,(\sec d)\left\{\frac{C_4(u+x)^2 + C_5(h + u\tan d)(u+x) - C_4(h + u\tan d)^2}{[(u+x)^2 + (h + u\tan d)^2]^2}\right\}du \qquad (16)$$

and, on integration over the cross section of the body,

$$\mathbf{F} = It\left(\frac{C_7 x + C_8 h}{x^2 + h^2}\right) \qquad (17)$$

where

$$\begin{aligned} C_7 &= 2C_4 \cos d + C_5 \sin d \\ C_8 &= -2C_4 \sin d + C_5 \cos d. \end{aligned} \qquad (18)$$

As was done in the case of a single dipole and in the case of an infinite line of dipoles, the expression for the field over a thin, dipping sheet (17) may be written in reduced form. Dividing numerator and denominator by h^2, and factoring out C_8, we have

$$\mathbf{F} = \frac{ItC_8}{h}\left(\frac{\Lambda_2\xi + 1}{\xi^2 + 1}\right) \qquad (19)$$

where $\xi = x/h$, and

$$\Lambda_2 = C_7/C_8 = \frac{\Lambda_1 \tan d + 2}{\Lambda_1 - 2\tan d} \qquad (20)$$

The reduced form is important, for it separates the parameters into two groups: the 'scale parameters,' which are effective in determining the size of the anomaly (here I, t, C_8 and h); and the 'shape parameters,' which are effective in determining 'type curves' representing the shape of the anomaly. Λ_2 is the parameter of a family of type curves obtained by plotting the function $(\Lambda_2\xi + 1)/(\xi^2 + 2)$ against ξ.

The thick, dipping polarized sheet (Fig. 4)—Take as an element a thin sheet whose top is at (u, h) and of width $du \sin d$. Then from (17) the force at Q is given by

$$d\mathbf{F} = I \sin d\left[\frac{C_7(x+u) + C_8 h}{(x+u)^2 + h^2}\right]du \qquad (21)$$

The force due to the whole sheet is, on integration across the sheet,

$$\mathbf{F} = I \sin d\left[\frac{C_7}{2}\ln(x^2 + h^2) + C_8 \tan^{-1} x/h\right]_2^1 \qquad (22)$$

where 1 and 2 represent the distances from $(-b, h)$ and (b, h) to Q. On substituting the limits, we have

$$\mathbf{F} = I \sin d\left[\frac{C_7}{2}\ln\frac{(x+b)^2 + h^2}{(x-b)^2 + h^2} + C_8\left(\tan^{-1}\frac{x+b}{h} - \tan^{-1}\frac{x-b}{h}\right)\right] \qquad (23)$$

The sloping step (Fig. 5)—Using a thin, horizontal sheet as an element and integrating, we obtain as the expression for the force at Q

$$\mathbf{F} = I \sin d\left[\frac{C_7}{2}\ln(x^2 + z^2) + C_8 \tan^{-1} x/z\right]_2^1 \qquad (24)$$

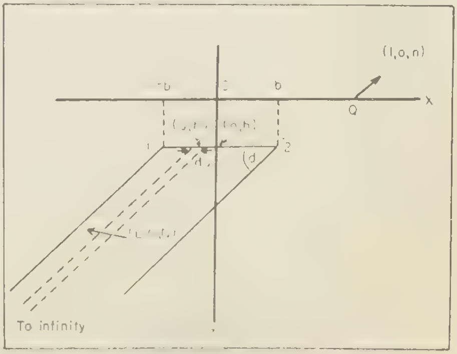


FIG. 4—Cross section of a thick, dipping, semi-infinite, polarized sheet, with element of integration.

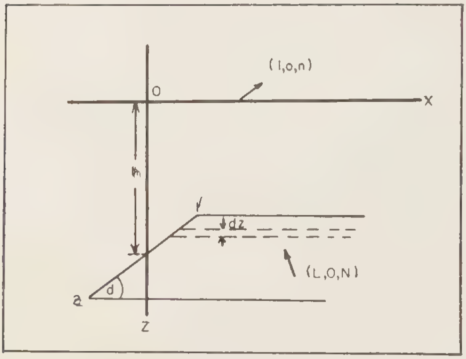


FIG. 5—Cross section of a sloping step, with element of integration.

where 1 and 2 represent distances from these points to Q, as shown in Figure 5. Applying these limits, we have

$$\mathbf{F} = I \sin d\left[\frac{C_7}{2} \ln \frac{(x - b \cot d)^2 + (h - b)^2}{(x + b \cot d)^2 + (h + b)^2} + C_8\left(\tan^{-1} \frac{x - b \cot d}{h - b} - \tan^{-1} \frac{x + b \cot d}{h + b}\right)\right] \qquad (25)$$

Comparing the above with previous derivations for the sloping step which treat only the special cases of measurement in the horizontal and the vertical directions *Heiland*, [1946, p. 397], we see that the more general form—suitable for use with aeromagnetic measurements or for treating arbitrary direction of polarization as an unknown to be solved for— is of no greater complexity than the forms arrived at in the more limited cases treated previously in the literature.

A system of equations of interpretation—Equations (22) and (24), giving the field over the thick, dipping, polarized sheet and the sloping step, are in the form

$$\mathbf{F} = c[F]_2^1 \qquad (26)$$

where $F = \alpha \tan^{-1} x/z + \beta \ln (x^2 + z^2)$ and c, α, and β are constants. (27)

The above equations were obtained by integrating, with respect to x, a function of the form of (17). We integrated the expression for an infinite line of dipoles (15) to obtain (17), the expression for a thin, dipping sheet. Thus D^nF, the nth derivative of F with respect to x, multiplied by c, will be the function representing the field or one of its derivatives over a thin sheet or a line of dipoles, depending on the body and the value of n.

Derivatives of F—If F is as given in (27), then

$$D^nF = D^{n-1}\left(\frac{\alpha h}{x^2 + h^2} + \frac{2\beta x}{x^2 + h^2}\right), \quad \text{with } n = 1, 2, \cdots \qquad (28)$$

Since we have [*Gibson*, 1931, p. 85]

$$D^r(x^2 + h^2)^{-1} = h^{-(r+2)}(-1)^r r! \sin^{r+1} \theta \sin (r + 1)\theta \qquad (29)$$

and

$$D^r(x/(x^2 + h^2)) = h^{-(r+1)}(-1)^r r! \sin^{r+1} \theta \cos (r + 1)\theta \qquad (30)$$

where $\cot \theta = x/h$, then substituting into (28) we have finally

$$D^nF = h^{-n}(-1)^{n-1}(n - 1)! \cdot \sin^n \theta(\alpha \sin n\theta + 2\beta \cos n\theta) \qquad (31)$$

where $n = 1, 2, \cdots$ This gives equations for use in forming the expressions for fields and derivatives over an infinite line of dipoles or a thin, dipping sheet. Application of the appropriate limits to (31) will produce expressions which may be used to form the equations for the fields and derivatives over the thick, dipping sheet or the sloping step.

In Table 1 expressions for D^nF for $n = 1$ to $n = 4$ are written.

TABLE 1—*Derivatives of F*

n	D^nF
1	$1/r^2(2\beta x + \alpha h)$
2	$-1/r^4(2\beta x + 2\alpha hx - 2\beta h^2)$
3	$2/r^6(2\beta x^3 + 3\alpha hx^2 - 6\beta h^2 x - \alpha h^3)$
4	$-6/r^8(2\beta x^4 + 4\alpha hx^3 - 12\beta h^2x^2 - 4\alpha h^3x + 2\beta h^4)$

TABLE 2—*Quantities represented by the derivatives in Table 1*

n	Line of dipoles $D^n\mathbf{F} = -\mu_L D^{n+2}F$ $\alpha = C_5; \beta = C_4$	Thin sheet $D^n\mathbf{F} = ItD^{n+1}F$ $\alpha = C_8; \beta = C_7/2$	Thick sheet or step (with limits applied) $D^n\mathbf{F} = I \sin d\, D^nF$ $\alpha = C_8; \beta = C_7/2$
0	$[D^0\mathbf{F} = \mathbf{F}]$ ...	...	Value of field
1	...	Value of field	Slope of profile
2	Value of field	Slope of profile	Second derivative (with respect to x or z)
3	Slope or profile	Second derivative (with respect to x or z)	Third derivative (with respect to x)
4	Second derivative (with respect to x or z)	Third derivative (with respect to x)	Fourth derivative (with respect to x or z)

As mentioned above, these derivatives refer to various quantities in connection with the profiles over the four bodies treated, depending on the body, the value of n, and whether or not limits are applied. These are summarized in Table 2.

Calculation of F and its derivatives—Values of the quantities treated above are often required in interpretation. Twelve different quantities are listed in Table 2, and any of these may be computed with the aid of the values of D^nF for the first four values of n. These are readily computed from the trigonometric form of the expression for the derivatives, and graphs of $\sin^n \theta \sin n\theta$ and $\sin^n \theta \cos n\theta$, from which these quantities may be obtained, are given in Figures 6 and 7. For lower values of n, these functions appear in expressions for the vertical and horizontal components of the fields over various bodies, and three of the curves for $n = 1$ and $n = 2$ appear in *Nettleton*'s Master Curves [1942, p. 299].

When $\xi = 0$, the absolute value of one of the functions is 1 and the other is zero for each value of n. Thus when $n = 1$ and $\xi = 0$,

$$\sin^n \theta \sin n\theta = 1 \quad \text{and} \quad \sin^n \theta \cos n\theta = 0,$$

and when $n = 2$ and $\xi = 0$,

$$\sin^n \theta \sin n\theta = 0 \quad \text{and} \quad \sin^n \theta \cos n\theta = -1.$$

Any of these functions which is zero for $\xi = 0$ will be denoted by f_0, and those whose absolute value is 1 at $\xi = 0$ will be denoted by f_1. Graphs of f_0 for various values of n are shown in Figure 6 and of f_1 in Figure 7. These are plotted against $|\xi|$, the absolute value of ξ, and represent f_0 and f_1 when preceded by the appropriate sign in the 'Table of Signs'. For example, f_0 for $n = 3$ is the negative of the values on the $n = 3$ curve on Figure 6 for $\xi < 0$.

As a numerical example, consider the horizontal line of dipoles for which, referring to Table 2 and (31),

$$\mathbf{F} = \frac{2\mu_L}{h^2}\left(C_4 f_1 + \frac{C_5}{2} f_0\right) \text{ for } n = 2 \qquad (32)$$

In calculating the value of $\mathbf{F}$ at $\xi = -0.23$, we see from the Table of Signs that f_1 is equal to the negative of the value on the $n = 2$ curve of Figure 7 for $|\xi| = 0.23$, and f_0 is the negative of the value on the $n = 2$ curve of Figure 6 at $|\xi| = 0.23$. Thus, for this value of ξ,

$$\mathbf{F} = (2\mu_L/h^2)(-0.856C_4 - 0.412C_5) \qquad (33)$$

Similarly, any of the quantities listed in Table 2 may be calculated.

Methods of solution—In the preceding sections, equations have been given for calculating the effect of arbitrary directions of polarization and measurement. These are in a form which leads to a simple and direct procedure for carrying out the interpretation of magnetic anomalies. It is now necessary to discuss the details of this procedure as applied to magnetic maps, where the field is given but the body is unknown, so that geological information may be extracted from the magnetic data.

Such maps give values of $\mathbf{F}$ over a horizontal

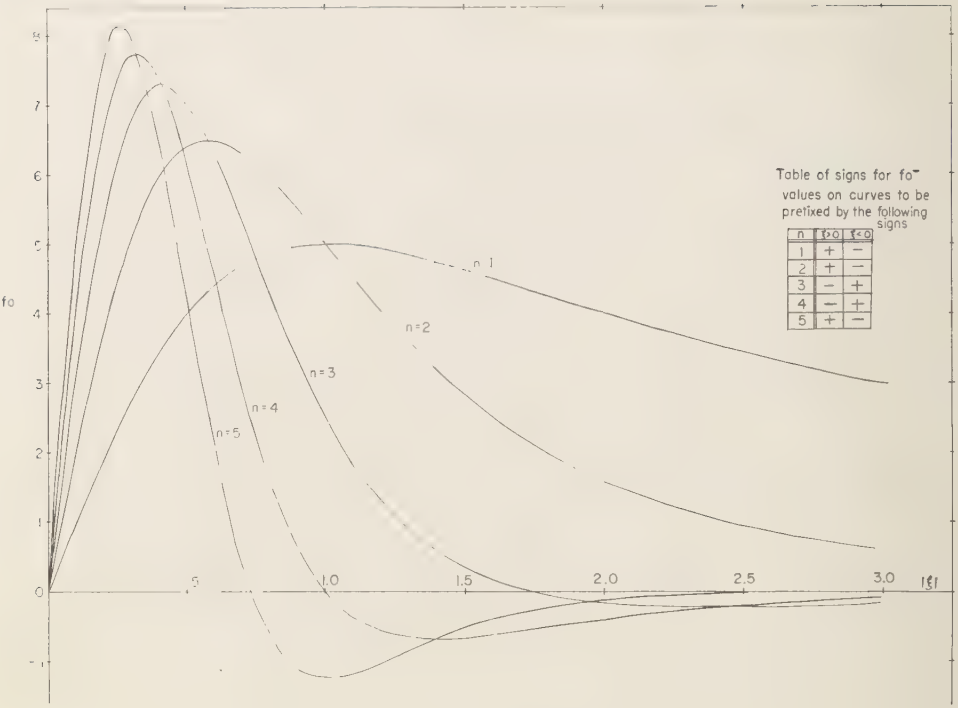


Table of signs for fo values on curves to be prefixed by the following signs

n	ξ>0	ξ<0
1	+	−
2	+	−
3	−	+
4	−	+
5	+	−

FIG. 6—Values of the function f_o.

plane. Interpretation is carried out by fitting the theoretical expressions for the fields due to particular bodies to the observed field, either from point to point in the plane of observation —as is done, for example, by *Vacquier* [1951]—or, as is more common, to profiles obtained by taking the values along some particular line. This latter, the method of profiles, will be adopted here.

Certain special points on the profile, because they are distinctive and easy to locate, and because the mathematical expressions for their positions are particularly simple, are especially suitable as a starting point in the interpretation. These are the maxima, the minima, and the inflection points. The first two are observable on the maps of **F** and the latter can be determined from these by numerical analysis.

Conditions for special points over the infinite line of dipoles and the thin, dipping sheet—These points occur for appropriate values of n when $D^nF = 0$, where F is as defined in (27). From (31) we see that this condition is satisfied when

$$\theta_n = 1/n \tan^{-1} (-2/\lambda) \qquad (34)$$

where θ_n is the value of θ corresponding to x_n, the position of the special point in question. $\lambda = \alpha/\beta$ and has the value C_5/C_4 or $2C_8/C_7$, depending on whether the body is a line of dipoles or a sheet.

Conditions for maxima and minima over a thick, dipping sheet—For the first derivative, $n = 1$ in Tables 1 and 2,

$$D\mathbf{F} = I \sin d \left[\frac{C_7 x + C_8 h}{x^2 + h^2} \right]_2^1$$

Inserting the limits, we find that

$$D\mathbf{F} = I \sin d \left[\frac{C_7(x + b) + C_8 h}{x + b)^2 + h^2} \right.$$

$$\left. - \frac{C_7(x - b) + C_8 h}{(x - b)^2 + h^2} \right] \qquad (35)$$

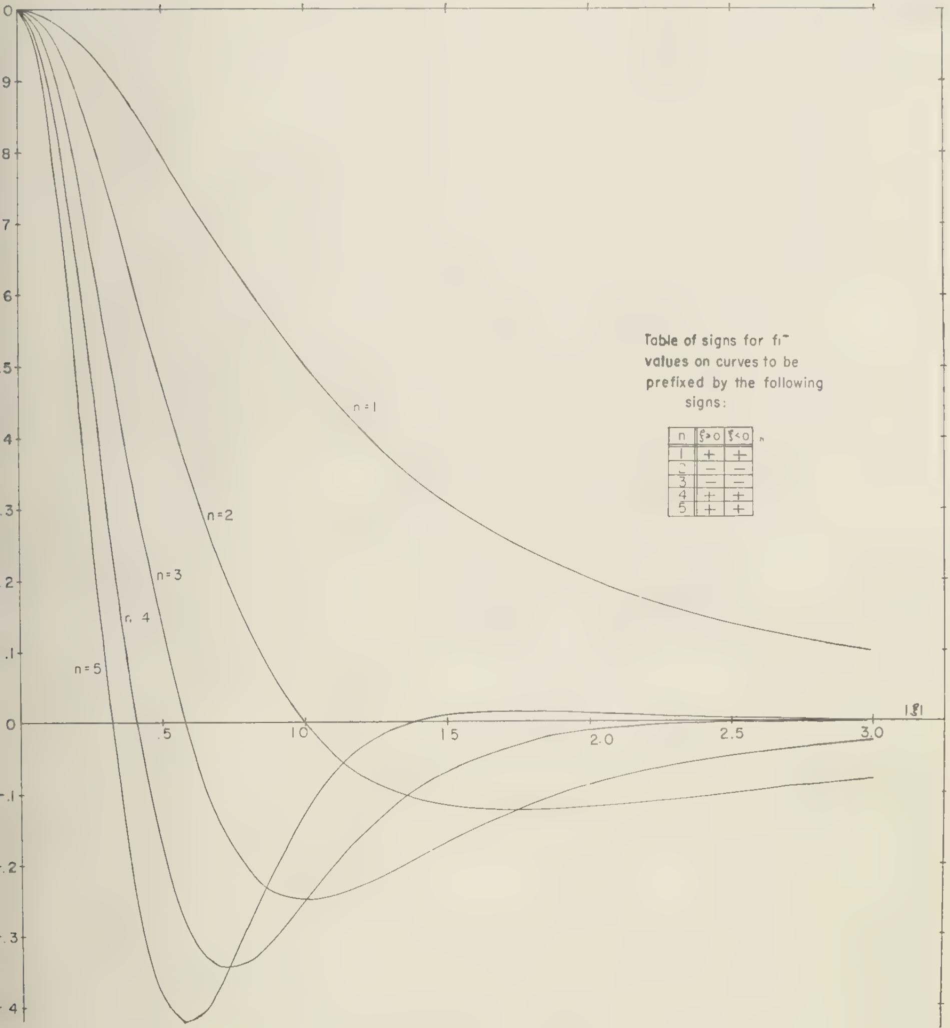


Table of signs for f_1 values on curves to be prefixed by the following signs:

n	$\xi > 0$	$\xi < 0$
1	+	+
2	−	−
3	−	−
4	+	+
5	+	+

Fig. 7—Values of the function f_1.

and equating this to zero, we obtain as a condition for maximum or minimum

$$\Lambda_2\xi^2 + 2\xi - \Lambda_2(1 + \beta^2) = 0 \qquad (36)$$

where

$$\xi = x/h, \beta = b/h \text{ and } \Lambda_2 = \frac{\Lambda_1 \tan d + 2}{\Lambda_1 - 2 \tan d},$$

$$\Lambda_1 = \frac{2(lN + nL)}{nN - lL} \qquad (37)$$

Conditions for maxima and minima over a sloping step—For the first derivative, $n = 1$ in Tables 1 and 2; from these, applying the appropriate limits,

$$DF = I \sin d\left[\frac{C_7(x + b) + C_8(h + b \cot d)}{(x + b)^2 + (h + b \cot d)^2} - \frac{C_7(x - b) + C_8(h - b \cot d)}{(x - b)^2 + (h - b \cot d)^2}\right] \qquad (38)$$

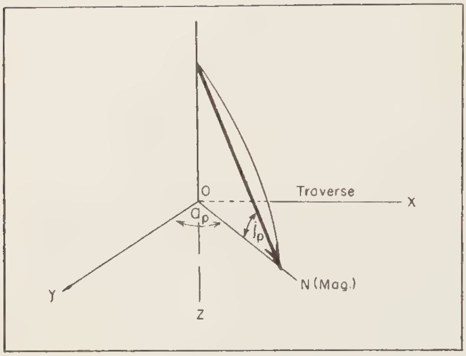


FIG. 8—Direction cosines of the polarization vector.

Equating to zero, we obtain

$$\xi^2 + \Lambda_1\xi + (\beta^2\Lambda_3 - 1) = 0 \qquad (39)$$

where Λ_1 and β are as in (37) and $\Lambda_3 = -(\Lambda_1 \cot d + \cot^2 d - 1)$. Thus, for example, if the dimensions and position of a step are known, then Λ_1 and hence L and N can be calculated from (39) if the positions of the maximum and the minimum on a profile over it are determined.

Calculation of parameters of bodies—The equations developed in the preceding sections may be used to calculate various unknown parameters of certain bodies from their magnetic profiles. Methods of doing this will be illustrated in the case of the infinite line of dipoles and the thin, dipping sheet.

For each of the bodies treated, the polarization function Λ_1, the dip d of any surfaces bounding the body, and their depths and magnetic moment per unit length along the strike combine in a way typical of the particular body to determine the size and shape of the anomaly. Among these, Λ_1 is a new quantity, arising from the treatment of directions of polarization and measurement as being completely arbitrary, and it requires a separate discussion.

The polarization function—This function, as defined in equation (15), for infinitely long, uniform bodies, depends upon the direction of polarization, the direction of measurement, and the angle at which the magnetic meridian cuts the strike of the body. It is most simply expressed in terms of direction cosines. Referring to (15), we see that since l and n depend upon the particular scheme of measurement adopte and are thus always known, the determinatic of Λ_1 from the anomaly leaves L and N as th only unknown quantities. The relation of thes to the direction of polarization is shown i Figure 8.

If i_p is the inclination and a_p the azimuth the polarization, then

$$L = \cos i_p \sin a_p$$

$$M = \cos i_p \cos a_p \qquad (4$$

$$N = \sin i_p$$

For infinitely long bodies of uniform cros section and polarization, $M = 0$. This is th case for all the bodies treated in this sectio L and N are not independent, for

$$L^2 + N^2 = 1 \qquad (4$$

The calculation of L and N from Λ_1 is facilitate by defining a quantity K given by

$$K = (\Lambda_1 n - 2l)/(\Lambda_1 l + 2_n) \qquad (4$$

Substitution into the expression for Λ_1 as give in (15) gives

$$N = l/(l + K^2)^{1/2}$$

and

$$L = K/(l + K^2)^{1/2} \qquad (4$$

Graphs of N and L against K are given (Figure 9 for use in calculating these quantities.

Calculation of the unknown parameters of body from the positions of the peak and the flankin inflection points of the associated anomaly—

1. *The infinite line of dipoles* (Fig. 2 and Ec (15)). The direction of polarization Λ_1, the dept h, and the magnetic moment per unit length c line μ_L are the unknown parameters. Any thre quantities depending upon these parameters ar sufficient to determine the unknowns. Referrin to Figure 10, we see that four such quantitie may be found from the positions of the pea and the flanking inflection points: $x_M - x_I' = w$ $x_M - x_I = w$, $\mathbf{F}_M - \mathbf{F}_I$, and $\mathbf{F}_M - \mathbf{F}_{I'}$ (wit primes indicating points or distances to th south of the peak). Referring to (34) and Table 1 we see that x_M is given by one of the solutions c

$$\tan 3\theta = -2/\Lambda_1 \qquad (44$$

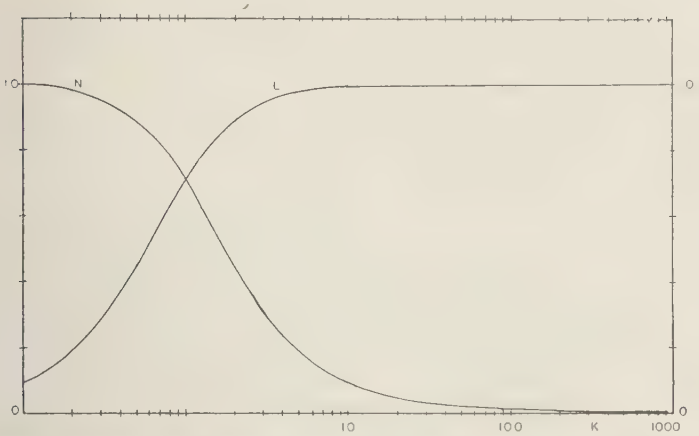


Fig. 9—Direction cosines of the polarization vector for bodies where $M = 0$; $N = 1/(1 + K^2)^{1/2}$, $L = K/(1 + K^2)^{1/2}$, $K = (\Lambda_1 n - 2l)/(2n + \Lambda_1 l)$; $L/N = \tan \varphi$, where φ is the dip of the polarization vector below the x-axis.

where $\cot \theta = x/h$; and x_I and $x'_{I'}$ are two of the solutions of

$$\tan 4\theta = -2/\Lambda_1 \tag{45}$$

Consequently, w/h and w'/h are functions of Λ_1 alone, as is w'/w. All these quantities may be calculated from the equations above and are shown in Figures 11 and 12, plotted against Λ_1. Since w/w' can be calculated from the profile, the corresponding value of Λ_1 can be obtained from the graph. It should be noted that w and w' are so related that the value of one for $-\Lambda_1$ is equal to that of the other for $+\Lambda_1$. If we know Λ_1, the graphs supply values of w/h or w'/h; since w and w' are known from the profile, h can then be calculated. From (15) it follows that, for an infinite line of dipoles,

$$\mathbf{F}_M - \mathbf{F}_I = 2\mu_L C_4/h^2\{(f_1)_M - (f_1)_I + \Lambda_1/2[(f_0)_M - (f_0)_I]\} \tag{46}$$

where the subscripts M and I represent values at the peak and the inflection point respectively of f_0 and f_1 as defined in connection with Tables 1 and 2. For a given direction of measurement,

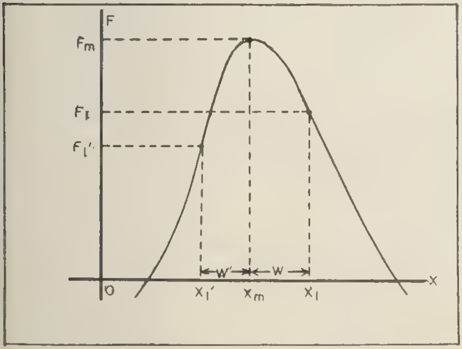


Fig. 10—Quantities related to the inflection points and peak on anomalies with a prominent maxium.

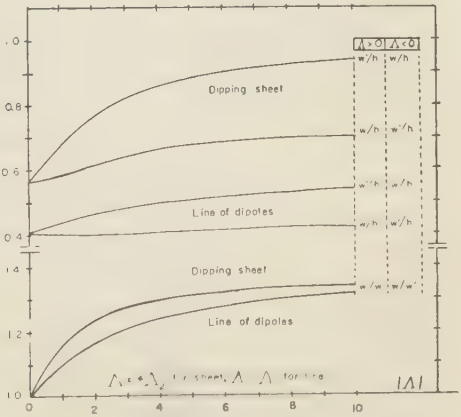


Fig. 11—Separation of peak and inflection points.

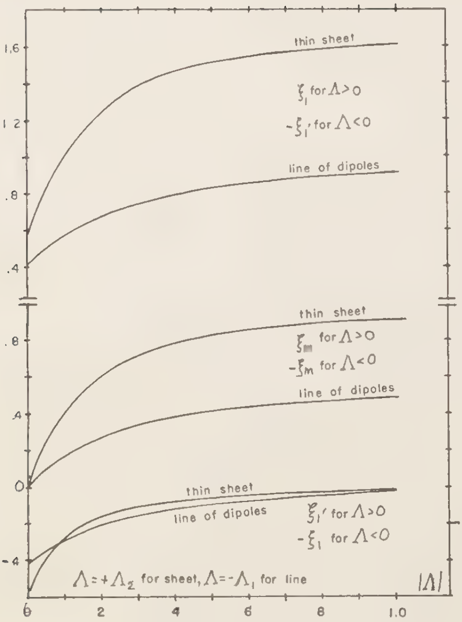


FIG. 12—Shift of peak and inflection points.

L and N can be found from Figure 9 and C_4 can be calculated from (13). This leaves μ_L as the only remaining unknown, which can then be found.

2. *The thin, dipping, polarized sheet* (see Fig. 3 and Eq. (17))—The direction of polarization, represented by Λ_1, the depth h, dip d, and the magnetic moment per unit surface area of the sheet It are the unknown parameters.

Exactly the same quantities as for the line of dipoles may be measured on the profile (Figure 10). For this body, referring to equation (34) and Table 2, we see that x_M is one of the solutions of

$$\tan 2\theta = -1/\Lambda_2, \quad \cot \theta = x/h \qquad (47)$$

and x_I and x_I' are two of the solutions of

$$\tan 3\theta = -1/\Lambda_2. \qquad (48)$$

These may be used to calculate values of w/h, w'/h and w/w', which are functions of Λ_2 alone. These quantities are all shown in Figure 11, plotted against Λ_2, and can be used in the same manner as the corresponding graphs for the line of dipoles to calculate h and Λ_2 for the dipping sheet, given a profile over that body.

From equation (19) it follows that for this body

$$\mathbf{F}_M - \mathbf{F}_I = ItC_8/h\{(f_1)_M - (f_1)_I + \Lambda_2[(f_0)_M - (f_0)_I]\} \qquad (49)$$

This may be used to calculate It.

From equation (20), the definition of Λ_2, we see that both the polarization function and the dip of the body combine to determine jointly the shape of the profile. Thus, d must be known from some other source of information, if Λ_1 and hence the direction of polarization is to be obtained from values of magnetic field strength over a sheet-like body.

Adjustments in the values obtained for the parameters—Once values of Λ_1, h, and μ_L have been obtained in the case of the infinite line of dipoles, or Λ_2, h, and It in the case of the thin, dipping sheet, equation (31), with appropriate values of n, α, and β, and the values from the graphs on Figures 6 and 7 may be used to calculate the values of $\mathbf{F}$ at successive points on the profile. With any of the schemes of curve fitting, the parameters may be adjusted to give an over-all best fit for the whole profile. Thus an adjusted set of parameters can be arrived at, which are not based on only the three points used as a start of the solution but on all the points of the profile.

Slopes at any point on the profile—For the line of dipoles, referring to (31) and Table 2, we see that

$$d\mathbf{F}/dx = -4\mu_L C_4/h^3(f_0 + \tfrac{1}{2}\Lambda_1 f_1),$$

with

$$n = 3 \qquad (50)$$

and for the thin, dipping sheet

$$d\mathbf{F}/dx = -ItC_8/h^2(f_0 + \Lambda_2 f_1),$$

with

$$n = 2 \qquad (51)$$

With values of the parameters of the body as calculated from the procedures outlined in the sections above, (50) and (51) may be used to calculate the slope of the profile at any point.

Example of calculation of parameters for an

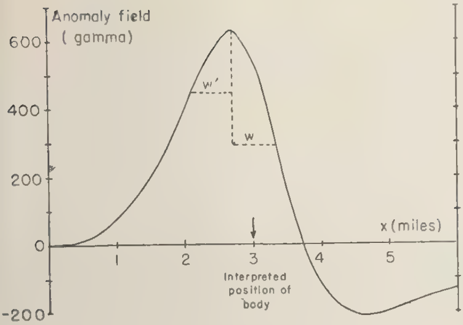


FIG. 13—Profile over a horizontal line of dipoles.

infinite line of dipoles—Figure 13 shows a theoretically calculated profile over an infinite line of dipoles. To illustrate the methods outlined above, let us begin with the profile and calculate the corresponding parameters of the body. Assume that the profile is over a body in high magnetic latitudes and that a magnetometer which measures along the direction of the total field is used; to a good approximation, $l = 0$ and $n = 1$.

Calculation of depth and polarization function—Assume that the inflection points have been located as shown on Figure 13. Then $w = 0.68$ mile, $w' = 0.60$ miles, $\mathbf{F}_M - \mathbf{F}_I = 335\gamma$, and $\mathbf{F}_M - \mathbf{F}_{I'} = 190\gamma$. Adopting the procedure described above, we find that $w/w' = 1.13$, and hence $\Lambda_1 = 1.50$. From this value of Λ_1, $w/h = 0.45$ and $w'/h = 0.40$. Knowing both w and w', we may make two estimates of h. Both of these give $h = 1.50$ miles. Also, from Figure 12, $\xi_M = -0.22$, $\xi_I = +0.23$, $\xi_{I'} = -0.62$, and thus $x_M = -0.33$ miles, $x_I = +0.35$ mile and $x_{I'} = -0.93$ mile. Referring to equation (32) and Figures 6 and 7, we find that at

$$\xi_M,\ \mathbf{F} = 2\mu_L C_4/h^2(-0.866 - 0.75(0.396))$$
$$= -2.326\mu_L C_4/h^2;$$

at

$$\xi_I,\ \mathbf{F} = 2\mu_L C_4/h^2(-0.856 + 0.75(0.412))$$
$$= -1.084\mu_L C_4/h^2,$$

and at

$$\xi_{I'},\ \mathbf{F} = 2\mu_L C_4/h^2(-0.324 - 0.75(0.649))$$
$$= -1.620\mu_L C_4/h^2.$$

Thus

$$\mathbf{F}_M - \mathbf{F}_I = -1.242\mu_L C_4/h^2 = +335\gamma,$$

and

$$\mathbf{F}_M - \mathbf{F}_{I'} = -0.706\mu_L C_4/h^2 = +190\gamma,$$

and in both cases,

$$2\mu_L C_4/h^2 = -539\gamma.$$

Calculation of the polarization—Λ_1 was found to be 1.50, and it was assumed that $l = 0$ and $n = 1$. Hence for equation (42), $K = \Lambda_1/2 = 0.75$. From Figure 9, $L = 0.60$ and $N = 0.80$. Thus $C_4 = lL - nN = -0.80$. The intensity of magnetization can now be computed from the value of $2\mu_L C_4/h^2$.

$$\mu_L = \frac{5.39 \times (1.5)^2 \times (1.609)^2 \times 10^7}{1.60}$$
$$= 1.98 \times 10^8 \text{ cgs units.}$$

If substituted into (32), these parameters give values of $\mathbf{F}$ sufficiently close to the observed profile for all values of x, showing that no adjustment of the parameters is required for this example. The values found above for the parameters of the body are the same as those originally used to calculate the profile.

Calculation of slopes at the inflection points—It is a useful check to compute these slopes from the calculated parameters and compare them with the observed values. The inflection points occur at $\xi_I = +0.23$ and $\xi_{I'} = -0.62$. Referring to equation (50) and Figures 6 and 7, we may write

$$d\mathbf{F}/dx = 719[-0.75(0.725) - 0.580]$$
$$= -810\gamma/\text{mile at } x_I$$

and

$$d\mathbf{F}/dx = 719[-0.75(-0.054) + 0.616]$$
$$= +470\gamma/\text{mile at } x_{I'}.$$

Example of calculation of parameters for a thin, dipping, polarized sheet—Figure 14 shows a theoretically calculated profile over a thin, dipping sheet with $\Lambda_2 = -0.85$ and $h = 0.50$ mile. These constants were chosen to produce a curve that is similar in size and shape to that

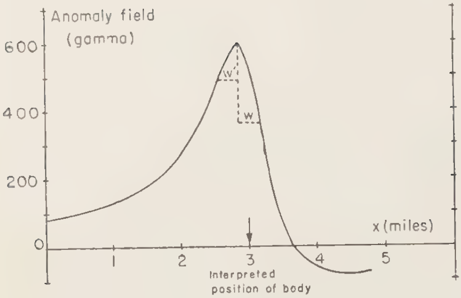


FIG. 14—Profile over a thin, dipping, polarized sheet.

for the line of dipoles, and it illustrates the similarity that can exist between profiles over these two types of bodies.

Assuming that the inflection points have been located as shown, we have $w = 0.32$ mile, $w' = 0.30$ mile, $\mathbf{F}_M - \mathbf{F}_I = 240\ \gamma$, and $\mathbf{F}_M - \mathbf{F}_{I'} = 105\ \gamma$. If we use a procedure similar to the one for the line of dipoles, we find that $w/w' = 1.07$; consequently $\Lambda_2 = -0.85$, giving $w/h = 0.65$ and $w'/h = 0.60$. Knowing both w and w' we may make two estimates of h. Both of these give $h = 0.50$ mile. Also, Figure 12 gives $\xi_M = -0.35$, $\xi_I = +0.30$, and $\xi_{I'} = -0.95$, leading to $x_M = -0.18$ mile, $x_I = 0.15$ mile, and $x_{I'} = -0.48$ mile.

Proceeding similarly to the formation of equation (32), we have

$$\mathbf{F} = ItC_8/h(\Lambda_2 f_0 + f_1) \quad \text{with} \quad n = 1 \qquad (52)$$

At

$$\xi_M,\ \mathbf{F} = ItC_8/h(+0.85(0.312) + 0.890) = 1.155\ ItC_8/h;$$

at

$$\xi_I,\ \mathbf{F} = ItC_8/h(-0.85(0.274) + 0.918) = 0.686\ ItC_8/h,$$

and

at

$$\xi_{I'},\ \mathbf{F} = ItC_8/h(+0.85(0.499) + 0.527) = 0.950\ ItC_8/h.$$

Thus

$$\mathbf{F}_M - \mathbf{F}_I = 0.469\ ItC_8/h, \quad \text{and} \quad \mathbf{F}_M - {}_{I'} = 0.205\ ItC_8/h,$$

and in both cases

$$ItC_8/h = 512\gamma.$$

Calculation of the dip and polarization—Referring to equations (19) and (20), we see that the shape of the profile depends only on h and Λ_2. Furthermore, Λ_2 is a function of both the polarization function and the dip. Thus in order to determine either of the last two quantities we must know the other from some independent source of information.

As an example, suppose that the dip of the sheet was suspected to be 65°S. The polarization function in terms of Λ_2 and d is given by

$$\Lambda_1 = \frac{2(\Lambda_2 \tan d + 1)}{\Lambda_2 - \tan d} \qquad (53)$$

thus

$$\Lambda_1 = \frac{2[-0.85(2.14) + 1]}{-0.85 - 2.14} = 0.55.$$

If the direction of measurement is such that $l = 0.31$ and $n = 0.95$, then $K = -0.05$, corresponding approximately to vertical polarization, for which we would have $N = 1$ and $L = 0$. In this case, $C_4 = -0.95$, $C_5 = -0.62$, and $C_8 = 1.46$. Thus It, the magnetic moment per unit area of the face of the sheet is given by

$$It = \frac{512 \times 0.50 \times 1.61}{1.46} = 282 \text{ cgs units.}$$

As in the case of the infinite line of dipoles, readjustment of the values obtained for the parameters proves to be unnecessary, since calculated values of **F** fall sufficiently close to the original profile. The values found for the parameters are the same as those originally used to calculate the profile.

Calculation of slopes at the inflection points—Referring to equation (51), we see that if $\xi = \xi_I = 0.30$,

$$d\mathbf{F}/dx = -1024[-0.85(-0.766) + 0.504] = -1185\ \gamma/\text{mile}.$$

For $\xi = \xi_{I'} = -0.95$,

$$dF/dx = -1024[-0.85(-0.026) - 0.525] = 515\ \gamma/\text{mile}.$$

These slopes are observed on the profile.

Summary and conclusions—Expressions for the magnetic force and its derivatives over a number of uniformly magnetized bodies have been derived for the case where both the directions of polarization and of measurement are arbitrary. This provides a greater generality in the equations for interpreting magnetic anomalies than in the forms previously published. In the present treatment the equations may be applied equally well to the special cases of vertical or horizontal directions of measurement, or of total field as measured by the nuclear free-precession magnetometer or by the type of fluxgate magnetometer used in present-day aeromagnetic surveying. In addition, any arbitrary direction of polarization may be incorporated into the calculation of the field or may be found as an unknown parameter.

For each of the bodies treated—the single dipole, the infinite line of dipoles, the thin, dipping sheet, the thick, dipping sheet, and the sloping step—expressions for the field and its derivatives have been expressed in a reduced form, which separates those parameters of the body which determine the shape of the anomaly from those which determine its size. Since inclusion of the direction of polarization gives a more complete set of parameters than those given in previous treatments, it is possible to make a more thorough assessment of the geological information obtainable from an analysis of the shape and size of magnetic anomalies.

It is concluded that from single profiles the depth, pole strength, and direction of polarization of a single dipole or a horizontal line of dipoles can be determined. For dipping sheets or a sloping step the depth, the pole strength per unit surface area, and a polarization function combining the direction of polarization and the dip of the inclined faces can be determined.

The equations for fields and their derivatives over infinite lines of dipoles, dipping sheets, and sloping steps are expressed as derivatives of a single function, and a general expression for derivatives of any order is given. Thus a single set of equations can be used to compute values of the fields and their derivatives over these bodies. Graphs are presented as an aid to their computation.

Points at which fields or their derivatives are zero are given special attention as the starting points for the analysis of the anomaly. Equations and graphs are given which show the effect of changing directions of measurement and polarization on the position of maxima, minima, and inflection points.

Acknowledgment—The author wishes to thank J. A. Jacobs of the University of British Columbia for critical discussion and review of the work at all stages. During the later stages of the work, the author was the recipient of a Canadian Kodak Fellowship.

References

Blackett, P. M. S., Lectures on Rock Magnetism, Weizmann Science Press, Jerusalem, Israel, 131 pp., 1956.

Gibson, G. A., *Advanced Calculus,* MacMillan, London, 503 pp., 1931.

Heiland, C. A., *Geophysical Exploration,* Prentice-Hall, New York, 1013 pp., 1946.

Jeans, J. H., *The Mathematical Theory of Electricity and Magnetism,* Cambridge University Press, 645 pp., 1948.

Mikov, D. S., Determination of the direction of polarization of disturbing bodies from the data of the magnetic survey (in Russian), *Izvest. Akad. Nauk SSSR, Ser. Geofiz.,* no. 5, pp. 418–423, 1953.

Nettleton, L. L., Gravity and magnetic calculations, *Geophysics, 7,* 293–303, 1942.

Sutton, D. J., and W. G. Mumme, The effect of remanent magnetization on aeromagnetic interpretation, *Australian J. Phys., 10,* 547–557, 1957.

Vacquier, V., Interpretation of aeromagnetic maps, *Geol. Soc. Am. Mem., 47,* 1–151, 1951.

Voskoboynikov, G. M., On the direction of magnetization of disturbing bodies in magnetic surveys (in Russian), *Izvest. Akad. Nauk. SSSR, Ser. Geofiz.,* no. 5, 483–485, 1955.

Yüngül, S., Prospecting for chromite with the gravimeter and magnetometer over rugged topography in east Turkey, *Geophysics, 21,* 433-454, 1956.

(Manuscript received November 10, 1958; revised July 24, 1959.)

Seismicity of the West African Rift Valley

J. Cl. De Bremaecker*

Institut pour la Recherche Scientifique en Afrique Centrale (I.R.S.A.C.) Bukavu, Belgian Congo

Abstract—All the epicenters determined in the central part of the West African Rift Valley up to the middle of 1958 have been plotted on a map. Most of them are on the faults which border the Rift; a few are on faults which crosscut it. The eastern Virunga 'extinct' volcanoes show a fairly strong seismic activity. The most important discovery is that of a transverse zone stretching west from Lake Kivu to the Congo River (450 km); extinct or active volcanoes are located at the intersection of this zone with the Rift Valley.

Diagrams show the amount of seismic energy liberated per year within 500 km of Lwiro. The mean value is 3.5×10^{20} ergs/year, which is about 0.03 per cent of that of the earth as a whole.

Introduction—The part of Africa that lies south of the Sahara is a very stable land mass, where a zone of mild seismic activity roughly follows the Rift Valley [*Gutenberg and Richter*, 1954].

In this paper we are concerned only with the area between 3°N and 6°S and 24°E and 32°E. The same area was studied by *Sutton and Berg* [1958].

Until 1953 the only two sources of information about the seismicity of this area were macroseismic observations and epicentral determinations of distant shocks. The former are usually single observations [*Cahen*, 1955], from which only very general conclusions can be drawn; the latter are hampered by the paucity of strong shocks, the distance to reliable seismographic stations, and the poor azimuthal distribution of these stations. The information antedating 1954 [*Gutenberg and Richter*, 1954] is summarized in Table 1.

It is worth remarking that shocks 1 and 3 are clearly outside the Rift Valley and shocks 2 and 4 are about 100 km from it. As will be seen later, this is not an abnormal occurrence.

Present situation—From 1953 to 1957 the Institute for Scientific Research in Central Africa (I. R. S. A. C.) installed four seismographic stations in the area: Lwiro, Uvira, Astrida, and Rumangabo, the latter in collaboration with the National Parks (Fig. 1). Each station is equipped with a Benioff vertical variable-reluctance seismometer which drives a 0.25-sec (Z) and a 25-sec (X) galvanometer and two Wood-Anderson instruments [*De Bremaecker*, 1955]. The Lwiro station has gradually been expanded to include Benioff horizontal seismographs and a set of long-period instruments on loan from Lamont Geological Observatory. The four stations form a diamond extending north-south, the sides and small diagonal of which measure about 120 km each.

Method—Since 1956 it has been possible to determine the epicenters of most well-recorded shocks, but the precision is naturally low for shocks more than 300 km away from the center of the network, especially to the north or to the south. The epicenters were determined with the help of the *Jeffreys-Bullen* tables [1948], because a uniform method of reduction is very desirable. It was realized that these tables might need corrections, but the errors involved are in any case relatively small.

The determinations were made for all shocks of magnitude ≥ 2. As the minimum magnitude increases with increasing epicentral distance, a fictitious crowding of shocks occurs near the stations. The records of the Benioff X and of the Wood-Andersons were compared in order to take advantage of the higher magnification of the former. The magnitude obtained from X was found to be 1.0 greater than the real value. A comparison of Z and X showed that 2.3 should

[*] Present address: Department of Geology, The Rice Institute, Houston, Texas.

TABLE 1—*Epicenters determined prior to 1954*

Shock	Date	Latitude	Longitude	Place	Magnitude
1	July 26, 1929	2.5°S	24.5°E	180 km NW of Kindu	5½*
2	March 18, 1945	0°	32°E	NW of Lake Victoria	6
3	Sept. 11, 1946	1.5°S	26.5°E	Transverse zone	5¾
4	Jan. 31, 1952	4°S	30.5°E	S. border of Urundi	6¼

*Class *d* of Gutenberg and Richter.

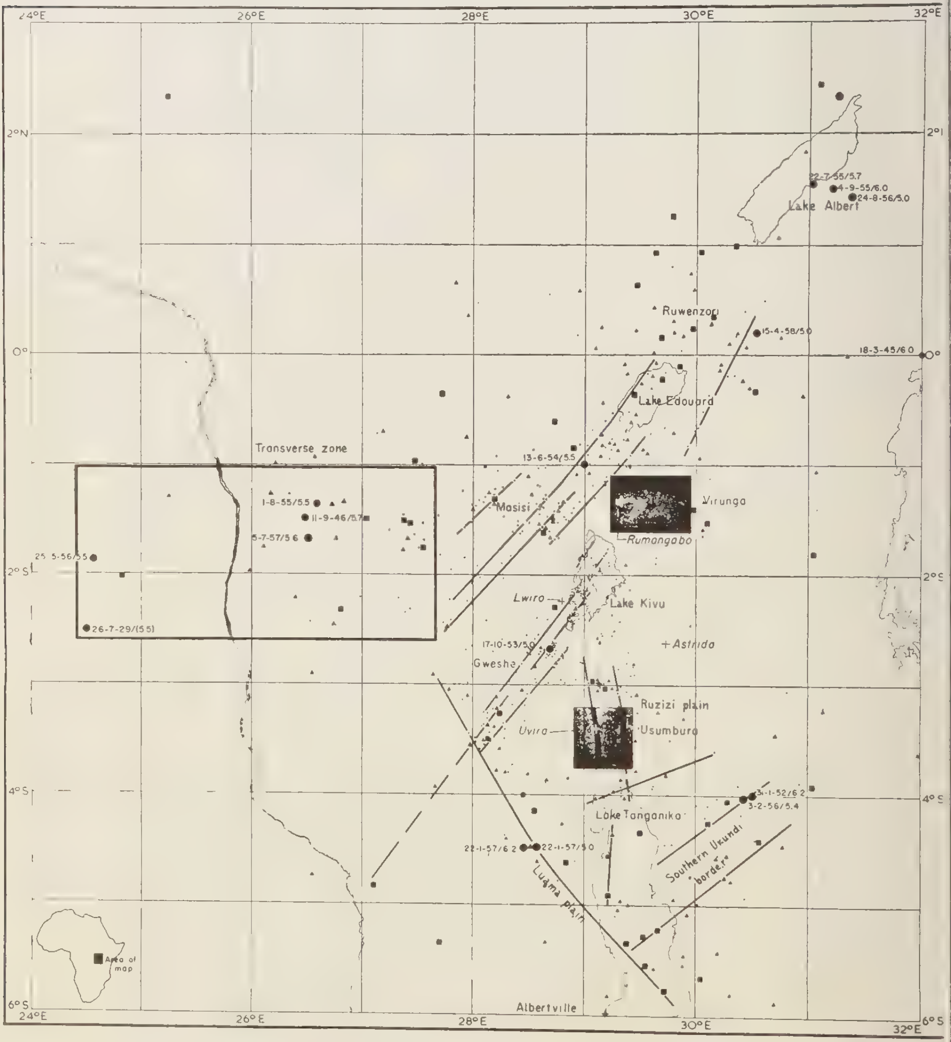


FIG. 1—Epicentral map of the area. Dots, shocks of $M < 3$; triangles, $3 \leq M < 4$; squares, $4 < M$ ≤ 5; circles, $M \geq 5$. For the latter the dates are given in the sequence: date-month-year/magnitud

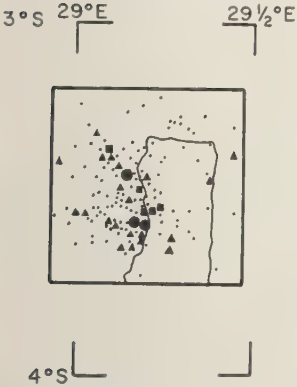


FIG. 2—Epicentral map of the Uvira-Usumbura area, at the northern end of Lake Tanganyika.

be subtracted from the magnitudes obtained from Z to obtain the real value. Because of obvious differences in instruments, these corrections are only approximate, but routine comparisons suggest that the magnitudes are generally correct to about ¼ *M*.

Epicenters—All epicenters determined prior to June 30, 1958, are shown in Figure 1; the details in the two black areas are shown in Figures 2 and 3. The faults are shown by heavy lines; they are drawn on alignments of epicenters, but the geomorphology of the area was also considered. It is not believed that every fault shown is continuous, but the pattern is thought to be useful as a first approximation.

Faults of the Rift or directly related to it—Most epicenters are clearly located along the faults which border the sides of the Rift or on their continuations; most faults coincide with the ones drawn by *Cahen* [1952]. It is immediately apparent that many shocks are quite distant from the Rift proper. If there is a similarity of structure between the Mid-Atlantic Ridge and the Rifts [*Ewing and Heezen,* 1956], this fact may be of interest, as it helps to explain the scattering of shocks on either side of the central part of the ridge.

From north to south the regions are

1. *Lake Albert area.* This region is too far from the stations for good determinations; it is one of the most active areas, where minor damage is frequent, contrary to *Willis'* statement [1936, p. 314].

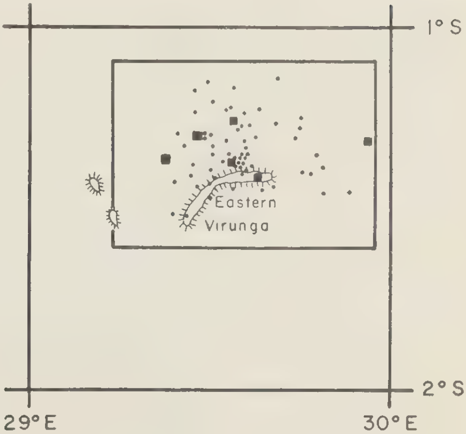


FIG. 3—Epicentral map of the eastern Virunga area. Shocks of $M < 3$ not indicated; dots, $M < 4$; squares, $4 \leq M < 5$.

2. *The Ruwenzori area.* This region is also too distant from the stations for definite comment; it is an area of moderate activity.

3. *Lake Edward area.* The shocks are mostly on the west side of the lake, or parallel to it but about 30 km to the east.

4. *The area southwest of Lake Edward* (NE of 1°S, 29°E). The alignments follow the ones in zone 3. There appear to be many faults, only a few of which have been shown.

5. *The Masisi area* (between 1°S and 2°S and 28°E and 29°E). This region continues the preceding ones but the faults are not geologically known. Some faults may go as far at 2.5°S.

6. *The Lake Kivu area.* Many parallel faults, still striking in the same direction, seem to be largely responsible for the topography. One of them separates the northern part of Idjwi Island (in Lake Kivu), which is flat, from the southern part, which is mountainous. The extinct volcanoes Kahusi and Biega, southwest of Lake Kivu, are not seismically active.

7. *The Gweshe.* This region continues the preceding one but the throw of the faults is reversed [*Cahen,* 1952].

8. *The Ruzizi plain.* The faults here strike N-S.

9. *The Uvira-Usumbura area.* This area is one of the most active at present and one in which minor damage is frequent; the faults also strike N-S (Fig. 2).

10. *Lake Tanganyika between Uvira and Albertville.* The precision of the epicenters decreases rapidly. Fairly strong shocks are occasionally felt in Kigoma and Albertville. In Kigoma a nearby earthquake caused a small scale tsunami on January 22, 1957.

11. *The southern border of Urundi.* The epicenters are often poorly determined in this area; the reason is unknown. One of the two strongest shocks of the area took place there, but because of sparse population no damage occurred.

12. *The Luama plain.* The fault is hypothetical, but it is shown by *Cahen* [1952]. It is suggested that the northern side of the Luama plain is bounded by a fault. The shock of magnitude 6.2 which occurred in this area on January 22, 1957, accounts for most of the energy released during that year.

Other areas—

1. *The Virunga volcanoes area* (Fig. 3). Most of these shocks are in the Rift Valley north of Lake Kivu but are presumably volcanic instead of tectonic. Only the shocks of magnitude ≥ 3 are shown. There are practically no shocks from the western Virunga group (active volcanoes).

The activity in the eastern area lasted from about June 18 to November 8, 1957, with irregularly alternating quiet and active periods. The eruption of the Mugogo [*Verhaege,* 1958] took place on August 1 and lasted 42 hours. Most of the shocks are on the northern flanks of the volcanoes.

2. *Transverse zone.* A seismically active zone stretches west from Lake Kivu for about 450 km. It measures about 150 km from north to south. Two of the four shocks mentioned in *Gutenberg and Richter* [1954] are located in this area. Moreover, a native town at 1.5°S, 26.5°E is called Kima, which means earthquake in the local dialect. The earthquakes sometimes cause damage as far as 100 km away. There is no sign that these shocks are deeper than normal.

It is a curious fact that this active zone has no topographic expression: in general the ground slopes gradually away from the mountains bordering the Rift and toward the Congo Basin. It might, however, have a very important geological expression: the extinct and active volcanoes (Virunga and South Kivu extinct volcanoes) are at the junction of this zone with the Rift Valley. We hope that gravimetric traverses across this zone will soon be made; the Rift itself is now generally considered as limited by normal faults [E. C. Bullard, personal communication, 1957; *Cahen,* 1952; *Goguel,* 1957; *Heiskanen and Vening Meinesz,* 1958]. For a different opinion see *McConnell* [1951, 1959].

Scattered shocks—Among the scattered shocks some may be on faults which have not yet been recognized and others may be similar to those occurring even in very stable land masses everywhere.

Discussion—The above results do not contradict those of *Sutton and Berg* [1958], except possibly in the transverse zone and in the southern Urundi area. It should be noted, however, that these authors did not determine the strike of the faults but took them as closely parallel as possible to the ones given by *Cahen* [1952], yet in agreement with seismological data.

In view of the present data it seems impossible to find a simple system of stresses that could produce all the directions of faulting which are shown. It is true that many faults strike ENE and that one strikes NW, but those north of Lake Tanganyika which strike due north cannot be fitted into any scheme. It thus appears that further discussion should be postponed until much more data are available.

Energy—In this section we have used only shocks within 500 km of Lwiro and of magnitude ≥ 3.2, which is the smallest shock recorded at this distance. This ensures the internal consistency of the data.

The energies were computed using the formula

$$\log E = 9.1 + 1.75M + \log (9 - M)$$

[*Gutenberg and Richter,* 1956, Eq. 21]. It is realized that this formula is not final, but the discussions center more on the constant term than on the linear one.

Figure 4 shows the sum of the energies by periods of 10 days. It is worth remarking that in 1955 more than 95 per cent of the energy was liberated in only 45 days. Table 2 shows the energy release per year.

The mean yearly value of the energy released is 3.0×10^{20} ergs; the mean value for the earth is roughly 10^{24} ergs (*Gutenberg and*

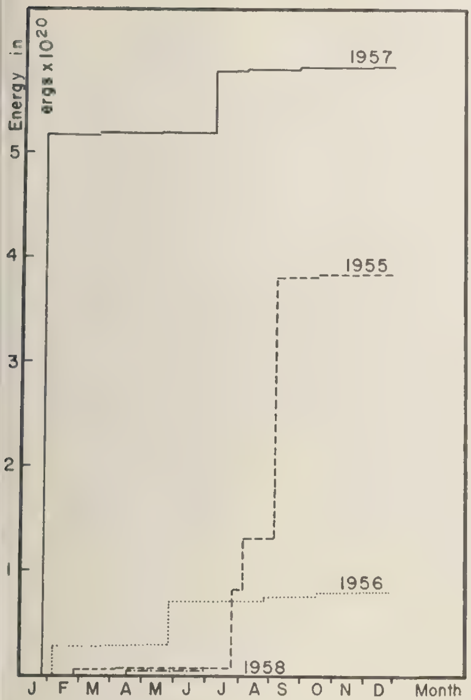


FIG. 4—Diagram giving the sum of the seismic energies released in periods of ten days for 1955, 1956, 1957, and the first half of 1958.

Richter [1954, p. 21] adjusted for the change in the constant term of the equation.) The energy released in the area is thus about 0.03 per cent of that for the earth. As the area considered is about 0.14 per cent of that of the earth, the mean value is almost 5 times more than the value in this area, even though it is one of the most active in Africa.

On the other hand, the mean value of the heat flux is 1.2×10^{-6} cal/cm^2 sec [*Birch*, 1954] or about 10^{25} ergs/yr in the area considered. This is about 30,000 times more than the seismic energy. There are no heat-flow measurements here, but the existence of active volcanoes and thermal springs suggest that the heat flow is higher than normal.

TABLE 2—*Energy release per year within 500 km from Lwiro*

Year	Energy in 10^{17} ergs
1955	3,860
1956	807
1957	5,850
1958 (first half)	50

Finally, it seems that the aftershocks of the earthquake of October 12, 1956, near Uvira ($M = 5$) were distributed roughly in an area of 300 km^2. Because of errors in some epicentral determinations it is possible that 100 km^2 would be more nearly correct. If we suppose, after *Utsu and Seki* [1954], that the 'aftershock area' is proportional to the magnitude, we find $\log A = M - 3$ (A in km^2). The above authors actually give $\log A = M + 6$ (A in cm^2), hence $\log A = M - 4$ (A in km^2). The wide differences in the magnitudes, in the instruments, and in the geological conditions are amply sufficient to explain this difference. It is, indeed, so small as to suggest that *Tsuboi's* idea [1956] of a variable earthquake volume is the more probable one, at least when a whole area is being investigated.

Conclusions—The most important result is the discovery of a transverse zone of activity stretching west of Lake Kivu for more than 400 km. Active or extinct volcanoes are located at the junction of this zone with the Rift. The other shocks are related to the Rift, generally clearly so.

Acknowledgments—Much of the data used have been gathered over the years by E. Berg, G. H. Sutton, and T. Janssen, to whom I am very grateful.

REFERENCES

BIRCH, FRANCIS, The present state of geothermal investigations, *Geophysics, 19,* 645–659, 1954.

CAHEN, LUCIEN, *Esquisse tectonique du Congo Belge et du Ruanda-Urundi,* Ministère des Colonies, Comm. Géol., 1952.

CAHEN, LUCIEN, *Géologie du Congo Belge,* Vaillant-Carmanne, Liège, 577 pp., 1954.

CAHEN, LUCIEN, *Bibliographie géologique du Congo Belge et du Ruanda-Urundi, 1–5,* Musée Royal du Congo Belge, 1955.

DE BREMAECKER, J. CL., Réalisations et programme de l' I.R.S.A.C. en seismologie, *Acad. roy. sci. coloniales, Brussels, Bull., 1,* 643–664, 1955.

DE BREMAECKER, J. CL., Premières données séismologiques sur le Graben de l'Afrique Centrale, *Acad. roy. sci. coloniales, Brussels, Bull., 2,* 762–787, 1956.

EWING, M., AND B. C. HEEZEN, Some problems of Antarctic submarine geology: *Antarctica in the IGY, Geophys. Monograph 1,* 1956.

GOGUEL, JEAN, Gravimétrie et Fossé Rhénan, *Koninkl. Ned. Geol. Mijnb. Gen., Geol. Ser., Gedenkboek F.A. Vening Meinesz, 18,* 125–147, 1957.

GUTENBERG, BENO, AND C. F. RICHTER, *Seismicity of the Earth,* Princeton Univ. Press, 310 pp., 1954.

GUTENBERG, BENO, AND C. F. RICHTER, Earthquake magnitude, intensity, energy, and acceleration (Second Paper), *Bull. Seis. Soc. Am., 46,* 105–146, 1956.

HEISKANEN, W. A., AND F. A. VENING MEINESZ, *The Earth and Its Gravity Field,* McGraw-Hill, New York, 470 pp., 1958.

JEFFREYS, H., AND K. E. BULLEN, *Seismological Tables,* Brit. Assoc. Advance. Sci., London, 50 pp., 1948.

MCCONNELL, R. B., Rift and shield structure in East Africa, *Rept. 18th Sess. intern. geol. Congr. London,* 1948, pt. 14, 199–207, 1951.

MCCONNELL, R. B., Outline of the geology of the Ruwenzori Mountains, *Overseas Geol. and Mir eral Resources, 7,* 245–268, 1959.

SUTTON, G. H., AND EDUARD BERG, Seismologic studies of the western Rift Valley of Afric *Trans. Am. Geophys. Union, 39,* 474–481, 1958.

TSUBOI, CHUJI, Earthquake energy, earthquak volume, aftershock area, and strength of th earth's crust, *J. Phys. Earth, 4,* 63–66, 1956.

UTSU, T., AND A. SEKI, A relation between th area of aftershock region and the energy mainshock, *Zisin, 7,* 233, 1954 (in Japanese, cite by Tsuboi, 1956.)

VENING MEINESZ, F. A., Les graben Africains, re sultat de tension ou de compression dans l croûte terrestre, *Inst. roy. colonial belge, Bul. 21,* 539–552, 1950.

VERHAEGE, MARCEL, L'éruption du volcan Mugog au Kivu, *Compt. rend., 246,* 2917–2920, 1958.

WILLIS, BAILEY, *East African plateaus and ri valleys,* Carnegie Inst. Wash., Publ. No. 47 358 pp., 1936.

(Manuscript received June 8, 1959; presented a the Fortieth Annual Meeting, Washington, D. C May 5, 1959.)

Calculations on the Thermal History of the Earth[1]

GORDON J. F. MACDONALD

Institute of Geophysics, University of California
Los Angeles, California

Abstract—The possible thermal history of a spherically symmetric earth is studied by comparing numerical calculations of the development of temperature within a number of model earths with observations on the present thermal state of the earth. The pertinent observations are the following: (1) The average surface heat flow is 50 ergs/cm^2 sec. (2) The mantle of the earth is solid. (3) The electrical conductivity increases rapidly in the outer few hundred kilometers of the earth. It is shown that models which incorporate a wide range of initial conditions and distributions of radioactivity reproduce in a qualitative fashion the electrical conductivity of the earth. Models which have a radioactive content equal to an earth composed of chondritic meteorites reproduce the present surface heat flow to within a factor of 2. This is a consequence of the near coincidence of the present rate of heat production in a chondritic earth and the heat flux of the actual earth. The variability in heat flow among models is due to varying contributions of 'initial heat,' higher rate of radioactive heat production in the past, and the depth of burial of heat sources. The solid nature of the mantle provides a further qualitative restraint on possible earth models. However, a lack of experimental data on melting relations at high pressures precludes detailed conclusions.

The general features of temperature-depth curves are similar for many models in which energy is transmitted by radiation as well as by ordinary lattice conduction. The gradient of temperature is high near the surface but decreases within the earth as the effective conductivity increases with increasing temperature. The melting temperature is most closely approached or exceeded in the outer few hundred kilometers of the earth.

The variation of surface heat flow with time is examined. It is shown that the surface heat flow is constant in time for a wide range of models, provided that heat is transmitted solely by radiation and conduction.

INTRODUCTION

The earth is a thermal engine. Any theory of the origin of the principal features of the earth must, in the end, include assumptions or deductions for its thermal history. Since Kelvin's classic investigation several important studies, in which ordinary thermal conduction was assumed to be the principal mechanism for heat transport, have been devoted to the determination of the earth's thermal character. *Slichter* [1941] emphasized the long time-scale associated with the diffusion of heat in the earth and the large thermal inertia of the earth. *Allan* [1955] and *Jacobs and Allan* [1954] have carried out detailed numerical calculations, taking into account the previously neglected time-dependence of radioactive heat production.

In recent years there have been several significant experimental developments contributing to an understanding of the earth's thermal history. Minute quantities of radioactive elements in geologically important materials can be determined by means of neutron-activation analysis. *Birch* [1958] remarked that the rate of heat production of an earth composed of stony meteorites, calculated from data obtained by neutron-activation analysis, is equal, within a factor of 2, to the rate at which heat is being lost by the earth. The extensive measurement of heat flow at sea by the Scripps Institution of Oceanography has done much to reduce the uncertainty as to the total rate at which heat is being lost from the earth.

At high temperatures large amounts of heat may be transported by radiation. *Clark* [1956] first noted the importance of radiative transfer in problems relating to the earth's thermal history, and he has initiated extensive laboratory investigations into the radiative conductiv-

[1] Publication No. 153, Institute of Geophysics, University of California.

ity of silicate materials. *Lubimova* [1958], making use of an analogue computor, carried out the first calculations on the earth's thermal history which took into account radiative transport of energy.

A number of observations serve to limit possible models of the earth's thermal history. The flow of heat from the interior of the earth averages about 50 ergs/cm^2 sec over both oceans and continents. (The energy unit used throughout the paper is the erg; 1 calorie $= 4.18 \times 10^7$ ergs.) Large regional variations exist, but the impressive feature is the apparent equality of average continental and oceanic heat flows. Seismology establishes the fact that the outer 2900 km of the earth is solid and that the material immediately underlying this mantle is fluid. A more accurate estimate of the melting points within the earth is now possible because of laboratory studies of the melting relations of geologically important materials at high pressures. Finally, even in silicate materials, thermal conductivity and electrical conductivity are not unrelated. The general features of the variations of electrical conductivity within the earth are known. The electrical conductivity depends sensitively on the temperature distribution and can be used to limit estimates of possible distributions of temperature presently within the earth.

The present study is concerned with the following question: What models for the earth's thermal history can be constructed that are in agreement with the observations mentioned in the preceding paragraph if we assume that thermal conduction coupled with radiative transport is the principal mechanism for heat transport? Models are studied in which the rearrangement of materials within the earth is taken into account but there is no explicit treatment of convected heat. The ratio of heat transported by radiation and conduction to that transported by convection is $K/L\rho Cv_r$ where K is the total conductivity, including radiation; C is the heat capacity: v_r is the radial velocity; ρ is the density; and L is the scale length. Substituting numerical values, we see that velocities of 10^{-10} cm/sec are required for convection to dominate radiation and conduction. The major limitation of treating only conduction and radiation in dealing with the earth's thermal problems is emphasized by these small velocities. As will be shown, it is possible, assuming only conduction to construct models that explain the major thermal features of the earth. However, these models are not simple, and the difficulties associated with them suggest that a detailed study of convective heat transport would be most profitable.

The paper is divided into two parts. In Part I the observational data and experimental studies relevant to the problems of the thermal history of the earth are reviewed. In Part II the methods of calculations are presented. The various models that have been studied are described in detail and the results of the calculations are compared with the observational data reviewed in Part I.

Part I: Review of Thermal Data

Heat generation by long-lived radioactive isotopes—Radioactive isotopes that have produced significant amounts of heat throughout the earth's history are distinguished by two characteristics: (1) The product of the abundance of the isotope and the rate of heat generated is relatively large. (2) The half-life of the isotope is of the same order as the age of the earth. The known isotopes that meet the two requirements are U^{238}, U^{235}, Th^{232}, and K^{40}. Isotopes with shorter half-lives may have played an important role in the initial stages of the development of the earth.

Birch [1951, 1954a] discussed the energy released in the radioactive decay of U, Th, and K. The energy associated with the decay is determined by measurements of the energy of all the individual particles and the radiation released during the decay. A correction must be made for the energy of neutrinos, since this energy is not absorbed in the earth. The neutrino energy depends on the energy in the beta spectrum, and it is customary to subtract two thirds of the maximum beta energy.

The decay of uranium and thorium results in alpha particles, beta particles, and gamma radiation. The alpha particles account for more than 90 per cent of the energy, so a good estimate of the energy associated with the decay can be made from counts of the alpha particles. The decay of K^{40} produces significant beta and gamma radiation. The energy released per dis-

TABLE 1—*Decay constants and heat generation of abundant radioactive isotopes*

	Disintegration energy				Heat generation		
Isotope	Mev/atom	ergs/atom × 10^{-6}	Decay constant, 10^{-10}/year	Half-life, 10^9 years	joules/g year	ergs/g sec	Percentage of abundance
U^{238}	47.4	75.9	1.54	4.51	2.97	0.94	99.27
U^{235}	45.2	72.4	9.71	0.71	18.0	5.7	0.72
Th^{232}	39.8	63.7	0.499	13.9	0.82	0.26	100.0
K^{40}*	0.71	1.14	5.5	1.3	0.94	0.298	0.0119
K^{40}†	...	...	5.30	1.25	0.92	0.288	0.0119
Rb^{87}	0.044	0.070	0.139	50	6.7×10^{-3}	2.1×10^{-3}	27.8
U	...	...	...	...	3.07	0.97	...
K	...	...	...	...	1.13×10^{-4}	3.55×10^{-5}	...
Rb	...	...	...	...	1.9×10^{-4}	5.8×10^{-5}	...

**Birch* [1951].
†*Aldrich and Wetherill* [1958].

integration of single atoms of the radioactive species is listed in Table 1.

For the uranium and thorium series the number of moles of the daughter nuclei n_D, produced in a time t, is

$$n_D = (e^{\lambda t} - 1)n_P \qquad (1)$$

where n_P is the number of moles of the parent nuclei present and where

$$\lambda = \frac{\ln 2}{\text{half-life}} = \text{decay constant} \qquad (2)$$

The decay constant gives the relative proportion of atoms decaying per unit time. *Birch* [1951, 1954a] and *Aldrich and Wetherill* [1958] gave detailed reviews of the data concerning the decay constants of heat-producing radioactive isotopes.

The half-lives of uranium and thorium are usually determined by measurement of the specific alpha activity. The measurement of the half-life of U^{238} presents no major difficulties; the total activity is approximately equal to twice the specific activity of U^{238} because the daughter isotope U^{234} contributes an equal activity. The contribution of U^{235} is small. The half-life measurements on U^{235} are more uncertain. The principal difficulty is due to the interference from alpha particles of U^{238} and U^{234}. Aldrich and Wetherill estimated that the decay constant of U^{238} might be in error by as much as 2 per cent. The half-life of Th^{232} has been measured by direct determination of the specific alpha activity of natural thorium and also by measurement of the 2.62 Mev gammas from Th^{208} in equilibrium with parent Th^{232}. The decay constants and half-lives of uranium and thorium are listed in Table 1.

The decay of potassium is more complicated. There are two decay constants, one giving the rate of beta decay of K^{40} to Ca^{40} and the other determining the rate of electron capture to A^{40}. The principal uncertainty in the decay of K^{40} is the importance of electron capture directly to the ground state of A^{40}. Only an upper limit for this mode of decay can be given, and, unfortunately, this upper limit is not well determined. In Table 1 the estimates of Birch and of Aldrich and Wetherill are listed. In the following calculations Birch's values have been used. It should be emphasized that these may be in error by as much as 10 per cent.

The decay constants and the energy yields can be combined to obtain the rate at which heat is produced by the radioactive disintegrations. The rates of heat production for the various isotopes are listed in Table 1 together with the proportion of the radioactive species.

The appropriate constants for rubidium are also listed in Table 1. The radioactive isotope Rb^{87} is more abundant than K^{40} by a factor of 20, even though Rb is less abundant than K by about a factor of 150 in meteorites [*Edwards and Urey*, 1955]. The specific activity of Rb^{87} is very low and is made up mostly of low-

Table 2—*Age of chondrites by argon-potassium method* [*Geiss and Hess, 1958*]

Meteorite	K, percentage by weight	Age, 10^9 years
Beardsley	0.100 ± .002	4.3
Mocs	0.087 ± .003	4.3
Bjurbole	0.085 ± .002	4.3
Holbrook	0.088 ± .002	4.4
Richardton	0.083 ± .002	4.1
Marion	0.087 ± .002	4.1
St. Michel	0.091 ± .002	4.0

energy beta particles. Thus, though Rb^{87} has a long half-life and is a relatively abundant isotope, it is not now an important heat-producing element. 2.0×10^{10} years from now, however, heat produced by Rb^{87} will determine the earth's thermal character.

Age of the earth—Consistent estimates of the age of chondrites have been obtained by using different radioactive-decay schemes. The results of *Geiss and Hess* [1958] are listed in Table 2. Earlier, *Wasserberg and Hayden* [1955] obtained three values for chondrites ranging between 4.1 and 4.4×10^9 years. *Schumacher* [1956], using the rubidium-strontium method, found similar values. These values are in accord with an estimate by Patterson [1956] of 4.5×10^9 years for two chondrites based on the ratio of Pb^{207}/Pb^{206}. The potassium-argon and strontium-rubidium ages refer to the time of the last separation of the parent and daughter nuclides during the formation of the meteorites.

Data on the age of the earth are much less consistent. The age of the oldest rock gives a lower limit. *Gerling and Pukanov* [1958] reported measurements on rocks showing an age of 3.4×10^9 years. The ratios of the stable lead isotopes have been used in numerous attempts to estimate the age of the earth. These estimates depend critically on the particular model for the history of the lead and on the composition of meteoritic lead. Reviews of the relevant data and estimates of age are given by *Wilson and others* [1956] and *Aldrich and Wetherill* [1958]. These estimates suggest that a value for the age of the earth of 4.5×10^9 years is not unreasonable. For the purposes of the calculation on the thermal history, a value of 4.5×10^9 years is adopted as the age of the earth. This age is taken as the time of the initial formation of the earth. It should be pointed out that the data on the age of meteorites and on the age of the earth are also consistent with the interpretation that the material which makes up the earth was aggregated at any time greater than 3.4×10^9 years ago and less than 7 to 10×10^9 years ago, though the upper limits require a very large initial abundance of U^{235} [*Burbidge and others*, 1958].

Chemical composition of chondrites—An earth of chondritic composition has been assumed in the calculation. The justifications for such an assumption were reviewed by *MacDonald* [1959]. *Urey and Craig* [1953] carried out a detailed analysis of the data relating to the bulk chemical composition of stony meteorites. For a study of a thermal history of the earth a knowledge of the content of potassium, uranium, and thorium in chondrites is critical. The potassium content in chondrites is remarkably uniform (Table 2). *Edwards and Urey* [1955] obtained an average potassium content of about 8.0×10^{-4} g/g for chondrites. The application of the neutron-activation analysis to the detection of uranium and thorium in stony meteorites has shed new light on the thermal problems of the earth [*Reed*, 1959]. *Hamaguchi and others* [1957] determined the uranium content of a number of chondrites and obtained an average value of

Table 3—*Heat-producing elements in chondrites*

Meteorite	U, 10^{-8} g/g	Th, 10^{-8} g/g	K, 10^{-4} g/g
Forest City	1.03*	4.0-4.7‖	8.4**
Modoc	1.08*	4.5‖	
Richardton	1.21*		8.3¶
Holbrook	1.26*	9.0‖	8.8¶
Modoc	1.1†		
Cumberland Falls	1.0‡		
Akaba	0.8§		
Beardsley		4.3‖	10.0¶

**Hamaguchi and others* [1957].
†*Tilton* [1956].
‡*Davis* [1950].
§*Reasback and Mayne* [1955].
‖*Bate and others* [1957].
¶*Geiss and Hess* [1958].
***Edwards and Urey* [1955].

TABLE 4—*Heat production by chondrites*

Meteorite	Heat due to U^{238}, ergs/g year	Heat due to U^{235}, ergs/g year	Heat due to Th^{232}, ergs/g year	Heat due to K^{40}, ergs/g year	Total, ergs/g year
Forest City					
present	0.30	0.013	0.36	0.94	1.61
4.5 × 10^9 years ago	0.60	1.06	0.45	11.1	13.2
Modoc					
present	0.32	0.014	0.37	0.90	1.60
4.5 × 10^9 years ago	0.64	1.14	0.46	10.6	12.8
Richardton					
present	0.35	0.016	0.39*	0.93	1.69
4.5 × 10^9 years ago	0.70	1.30	0.49	11.0	13.5
Holbrook					
present	0.37	0.016	0.74	0.98	2.11
4.5 × 10^9 years ago	0.74	1.30	0.92	11.6	14.6
Beardsley					
present	0.31*	0.014*	0.35	1.12	1.79
4.5 × 10^9 years ago	0.62	1.14	0.44	13.2	15.4
Model Earth					
present	0.32	0.014	0.36	0.90	1.59
4.5 × 10^9 years ago	0.64	1.14	0.45	10.6	12.8

*Estimated on basis of Th/U = 4.

1.1×10^{-8} g/g. Individual values are listed in Table 3. Neutron-activation analyses of the thorium contents of a number of chondrites were reported by *Bate and others* [1957]. These values are listed in Table 3. It should be noted that the ratio of thorium to uranium is about 4, except in the case of the Holbrook meteorite. This ratio is consistent with observations of the crustal materials. Furthermore, *Marshall* [1957], in a study of the isotopic composition of common leads in crustal materials, estimated a thorium-to-uranium ratio of 4:1.

The uranium-thorium content of iron meteorites is at least 2 to 3 orders of magnitude less than that of chondrites. *Bate and others* [1958] found in two iron meteorites a range from 6×10^{-12} to 2×10^{-11} g/g of thorium. Reed, Hamaguchi, and Turkevich found, by neutron-activation analysis, uranium concentrations in five iron meteorites ranging from 3×10^{-11} to 1×10^{-10} g/g.

Concentrations of uranium and thorium in chondrites, as determined by the neutron-activation method, run about an order of magnitude less than the amounts obtained by chemical methods. The validity of neutron-activation analysis is still open to question. The fact that consistent results are obtained with different nuclear and chemical species argues for the validity of the method. The danger of peculiarities in resonance effects and in the chemical behavior of the species is always present. Since the values obtained from neutron-activation analysis form the basis of the calculations of the thermal history of the earth, it is important to emphasize that further revisions in the estimates of the radioactive content in meteorites may still lie ahead.

Heat production in chondrites—Estimates of the present rate of energy release in chondrites and the heat production of chondrites 4.5 × 10^9 years ago are given in Table 4. These estimates are based on the radioactive element content listed in Table 3. The total amount of heat produced by the five chondrites is remarkably uniform. At present about three fifths of the heat being produced is due to the disintegration of K^{40}, with uranium and thorium each contributing about one fifth of the total heat now being produced.

The rate of heat production at a time t years ago is $e^{\lambda t}$ times the present rate, where λ is the decay constant. Because of the shorter half-life of U^{235} and K^{40} the relative contribution to the heat production by these two isotopes was much greater 4.5 × 10^9 years ago. The total amount of heat which was produced in these chondrites at this earlier time was about eight times the present rate of heat production.

Table 4 lists the radioactive heat production corresponding to a content of 1.1 × 10^{-8} g/g of uranium, 4.4 × 10^{-8} g/g of thorium, and 8.0 × 10^{-4} g/g of potassium for the model earth which is later used in the calculations of the thermal history. It will be noted that these values are about equal to the radioactive heat production of the Forest City and the Modoc chondrites. The lower values are taken, since it is believed that the principal errors in the neutron activation analysis come from contamination of the meteorite. Contamination leads to higher values of the radioactive isotopes, since materials coming in contact with the meteorite undoubtedly have a higher concentration of radioactive isotopes than the meteorite has.

Heat production in crustal materials—The distribution of temperature in the outer few tens of kilometers of the earth depends on the content of radioactive elements in crustal rocks. A large number of analyses of surface materials have been made. The problem of estimating the radioactive content of the crust is that of assigning the proper weights to the various types of rocks. This problem is unsolved.

TABLE 5—*Average content of uranium and thorium in igneous rocks*

Type of rock	Number of samples	U $\times 10^{-8}$ g/g	Th $\times 10^{-8}$ g/g
Granites*	9	380	1030
Acidic† (range of averages for Canada and U. S.)	1257	380-400	1310-1350
Intermediate*	6	140	440
Intermediate†	297	230-300	930-1050
Basalts*	8	83	500
Basic Rocks*	27	95	380
Basic Lavas‡		60-110	
Hualalai Basalt§,‖	1	46-50	

**Evans and Goodman* [1941]
†*Senftle and Keevil* [1947]
‡*Adams* [1954]
§*Hamaguchi and others* [1957]
‖*Tilton* [1956]

Estimates of the uranium and thorium content of broad classes of rocks are listed in Table 5. In addition, the Hualalai basalt is listed separately, since its uranium content has been determined both by neutron activation and by mass spectrometry. Uranium and thorium tend to be concentrated in granites and to be less abundant in basalts. There is a general tendency for uranium and thorium to increase with the silicon content. The ratio of thorium to uranium averages about 4.

The uranium content of dunites is much smaller than that of basalts. The analytical difficulties are correspondingly greater. The range of values obtained by neutron activation, by mass spectrometry, and by wet analysis for radium are given in Table 6. The analyses by neutron activation may be more reliable because they are less affected by contamination There are as yet no data on the thorium content of dunites.

On the basis of the data listed in Tables 5 and 6, the heat produced by classes of rocks is shown in Table 7. Unlike meteorites, the principal heat producers in crustal materials are uranium and thorium. Uranium and thorium produce approximately equal amounts of heat and potassium produces about a third as much as either. If these estimates are in any way reliable, it would seem that uranium and thor-

TABLE 6—*Uranium content of dunites*

Dunite	U, 10^{-8} g/g	Analytical method
Twin Sisters*	0.10-0.12	neutron activation
Twin Sisters†	1.6	mass spectrometer
Twin Sisters‡	2.4	radium
Balsam Gap, N. C.‡	0.9-1.2	"
Dun Mtn.‡	0.6	"
Addie, N.C.‡	2.1	"
Webster, N. C.‡	0.9	"

**Hamaguchi and others* [1957]
†*Tilton* [1956]
‡*Davis and Hess* [1949]

Table 7—*Heat production by igneous rocks*

Type of rock	Heat produced by U, ergs/g year	Heat produced by Th, ergs/g year	Assumed content of K, 10^{-4} g/g	Heat produced by K, ergs/g year	Total, ergs/g year
Granites	117	84	300	34	235
Acidic	126	109	340	38	273
Intermediate	43	36	263	29	108
Intermediate	81	81	263	29	191
Basalts	25	41	57	6.4	72
Basic Lavas	26	28*	49	5.5	59
Hualalai Basalt	15	16*	56	6.3	37
Twin Sisters dunite (neutron activation)	0.034	0.036*	0.1	0.01	0.08
Dunites	0.42	0.44*	0.1	0.01	0.87

*Calculated on basis of Th/U = 4.

ium are much more concentrated in the crust than is potassium. The degree of differentiation of uranium and thorium is then much greater than that which has affected potassium. The data are scanty, but it appears that the heat production of dunite is a tenth or a hundredth that of basalts.

A unique estimate of the content of the radioactive elements in the crust is impossible. Table 8 lists the mass of the radioactive elements in the continental crust provided the crust is made up of the intermediate group of rocks listed in Table 5. The oceanic crust is supposed to be made up of the basaltic rocks. Under this assumption the total mass of radioactive elements in the crust is less than that of a chondritic model earth. The proportion of uranium and thorium in the crust may be as high as a third to two thirds of the total amount of uranium originally within a chondritic earth. The proportion of potassium is much less, on the order of one sixth. The values listed in Table 8 must be regarded as crude estimates; the total mass of the radioactive elements in the crust may be somewhat greater or may be a great deal smaller. The principal uncertainty is in the assignment of weights to the various types of rocks. But there is additional uncertainty because many major groups of rocks have not yet been properly analyzed for uranium and thorium. This is particularly true for the metamorphic rocks. A revealing experiment would be the measurement of heat flow at the base of the continental crust. Such determinations would help to establish the radioactivity of the crust.

Heat flow measurements—A most important quantity in any discussion of the thermal character of the earth is the amount of heat that is presently escaping from the earth's interior. The outward flux of heat by conduction per unit area and per unit time is equal to the product of the thermal conductivity and the temperature gradient,

$$dQ/dt = -K(\partial T/\partial r) \qquad (3)$$

The observation of heat flow requires separate

Table 8—*Mass of heat-producing elements in the crust*

Region	Mass, g	Mass of U × 10^{-19} g	Mass of Th × 10^{-19} g	Mass of K × 10^{-23} g
Continental crust	1.76 × 10^{25}	2.5-4.6	7.7-17.0	4.6
Oceanic crust	0.62 × 10^{25}	0.3-0.6	1.2-3.1	3.5-3.8
Total crust	2.38 × 10^{25}	2.8-5.2	8.9-20.1	8.1-8.4
Model earth	5.98 × 10^{27}	6.6	26.0	48.0

measurements of thermal conductivity and of the temperature gradient.

Systematic studies of thermal conductivities and temperature gradients have been carried out only since 1939. Bullard, Birch, and their co-workers, together with Revelle, Maxwell, and Von Herzen of the Scripps Institution of Oceanography, are responsible for most of the reliable measurements on heat flow. *Birch* [1954a, 1954b] reviewed the measurements on heat flow. At that time about forty determinations had been made in land areas and another twenty-five had been carried out at sea. The land areas showed a variation, roughly, of from 20 to 120 ergs/cm² sec. The variation in the ocean measurements was somewhat greater, ranging from 15 to 140 ergs/cm² sec. The mean of these measurements indicated an average heat flow of about 50 ergs/cm² sec. Since that time the number of measurements at sea in the Pacific has more than doubled [*Von Herzen*, 1959]. The new measurements indicate that the mean heat flow through the Pacific is still about 50 ergs/cm² sec, but that there is a far wider range in individual values—from 10 to 370 ergs/cm² sec. A partial explanation of this greater range in oceanic values may be that it is far easier to avoid thermally active regions on land than it is at sea. However, there is no simple explanation of the rather widespread lower values that have been found in the Pacific.

Measurements of the heat flow, restricted on land to non-thermally active regions, give an estimate of the amount of heat brought to the earth's surface by conduction. *Birch* [1954a] estimated that the rate of heat brought to the surface by hot springs and volcanoes is less than 5 ergs/cm² sec.

Thermal conductivity of silicate materials—*Birch and Clark* [1940] demonstrated that the thermal conductivities of a wide variety of rocks show surprisingly small variations with composition. Recent measurements at room temperature [*Clark and Niblett*, 1956; *Birch*, 1950] further established the narrow variation of conductivity with composition. Some of the data relating to low-temperature conductivity of silicate materials are summarized in Table 9.

TABLE 9—*Thermal conductivity of rocks at moderate temperatures*

Type of rock	Number of samples	*K*, joules/cm sec °C at 30°C	 at 200°C	 at 400°C
Granite* and Quartz Monzonite	59	0.0330		
Quartz Diorite* Gneiss	17	0.0324		
Injection Schist* and Gneiss	41	0.0324		
Gneiss†	38	0.0272		
Limestone and Dolomite†	5	0.0359		
Granite‡	4	0.0298	0.0250	
Anorthosite‡	3	0.0177	0.0185	
Diabase‡	3	0.0220	0.0215	0.021
Gabbro‡	3	0.0206	0.0204	0.020
Dunite‡	3	0.0477	0.0340	

**Birch* [1950].
†*Clark and Niblett* [1956].
‡*Birch and Clark* [1940].

Birch and Clark measured conductivities up to 400°C and found that the conductivities of poor conductors, such as feldspar-rich material increased with temperature, whereas the con ductivities of most materials decrease as th temperature increases. Studies of dielectric soli at somewhat higher temperatures show that th conductivity decreases approximately inverse with the temperature in accordance with phono theory [*Peierls*, 1955]. No experimental studi have been carried out on the variation of con ductivity with pressure.

Lubimova [1958] suggested on the basis lattice-conduction theory that the conductivit increases with pressure. The theory of therm conductivity is difficult, and its applicability t silicate materials has not been adequately teste Furthermore, a test of the conclusion that la tice conductivity increases with pressure cann be made because of the lack of data relating t conductivity at high pressures.

At high temperatures energy may be tran ferred within a solid by radiation rather tha by lattice vibrations. *Clark* [1957a] shows th the contribution of radiation to thermal con ductivity is approximated by

$$K_R = 16n^2sT^3/3\epsilon \qquad (4)$$

In this expression n is the refractive index of the material, s is the Stefan-Boltzman constant, ϵ is the sum of the absorption and scattering coefficients averaged over all wavelengths, and T is the temperature in degrees Kelvin. The striking feature of the contribution of radiation to the conductivity is its strong temperature-dependence. The higher the temperature, the greater will be the energy transferred by radiation, provided that the variation of index of refraction and opacity do not overwhelm the T^3 dependence.

The opacity ϵ and the index of refraction are determined by the properties of material and in general will be functions of temperature and pressure. The variation of the index of refraction of a given material with pressure has not been investigated either experimentally or theoretically. *Clark* [1957a] and *Lawson and Jamieson* [1958] suggested that the pressure dependence can be estimated from the empirical relation determined by Gladstone and Dale. In this relation the index varies linearly with the density. Though the Gladstone-Dale relation is reasonably well satisfied by different minerals [*Larsen and Berman*, 1934], it is uncertain whether such a relation can be applied to a single material subjected to large changes in temperature and pressure. In the calculations discussed in Part II, a constant index of refraction, $n = 1.7$, is assumed. The mean index of forsterite is 1.65. The order of the error introduced into the calculations by neglecting the variation of the index of refraction with pressure is easily estimated. Using the Gladstone-Dale relation and taking the density of the base of the mantle as 5.7, we find that the index of refraction of the material at this level should be 2.2, provided that it is 1.7 at the top of the mantle. Since the radiative conductivity depends on the square of the index of refraction, we shall underestimate the radiative conductivity at the base of the mantle by a factor of $(2.2/1.7)^2 = 1.65$. In view of other uncertainties in estimating the conductivity of material within the earth, the uncertainty due to the variation of index of refraction is thought to be small.

The rate at which radiation is transferred through a solid depends on the frequency or wavelength of the radiation. Various mechanisms are known by which solids can absorb radiation, and these become important at different frequencies. Absorption due to the excitation of lattice vibrations by the radiation is strong in the infrared. This lattice absorption is relatively unimportant, since at high temperatures the energy density is low at these long wavelengths. Intrinsic absorption is due to the excitation of valence electrons to a conduction band across the fundamental energy gap. Intrinsic absorption is important in the ultraviolet for wavelengths of less than about 0.4 microns [*Runcorn and Tozer*, 1955]. The transparency of silicates to radiation is thus limited at long wavelengths by the infrared absorption due to lattice vibrations and at short wavelengths by the absorption in the ultraviolet due to the excitation of electrons to the conduction band.

The region of high transparency in silicates lies in the visible and near infrared. This is not true for semiconductors, such as silicon and germanium, which are opaque in the visible. Because of a lower energy gap the absorption edge in these materials lies at much longer wavelengths than in silicates. Within the visible and near infrared region of transparency two processes can lead to absorption. Characteristic absorption peaks are associated with the presence of transition elements. The energy levels of the unfilled electron shells are split by the crystalline field and certain transitions between these split levels are allowed. These transitions lead to characteristic absorption bands of the transition elements. In silicates the most important transition element is iron, which has a strong absorption peak at about 1 micron. Titanium, manganese, and other elements will also give rise to absorption bands. It is these bands that give color to the crystals.

As S. J. Clark (personal communication) has pointed out, it is the absorption between the peaks that determines the contribution of radiation to the thermal conductivity. If there is one perfectly transparent region, then the material has an infinite thermal conductivity. The general level of absorption between absorption bands limits the energy transported by radiation. This general absorption is primarily due to free electrons. In the classical theory free electrons will absorb at all wavelengths. The dependence of the opacity on electrical conductivity is given by

$$120\pi\sigma/n \tag{5}$$

where in general both the electrical conductivity σ and the index of refraction n vary with the frequency. *Clark* [1957] supposed that the conductivity at zero frequency was a sufficiently good approximation. In this case the variation of the opacity can be written as

$$\epsilon = \epsilon_0 + \frac{120\pi\sigma_0 e^{-E/2kT}}{n} \tag{6}$$

where ϵ_0 is the opacity at low temperatures and where the second term on the right-hand side of the equation is proportional to the electrical conductivity. *Lawson and Jamieson* [1958] criticized Clark's use of the zero-frequency approximation.

From the preceding discussion it is apparent that the region of transparency in the near infrared and visible (between the ultra-violet absorption edge and the infrared lattice absorption) can be closed either by the presence of absorption peaks due to transition elements or by a high level of general absorption due to free carriers—in particular, free electrons. Furthermore, it must be expected that changes in temperature and pressure will affect the nature of the region of transmission.

A few direct measurements of thermal conductivity at high temperatures suggest the importance of radiation [*McQuarrie,* 1954a, b]. Most estimates of the contribution of radiation to the conductivity are based on determination of the absorption spectra rather than on actual measurements [*Eitel,* 1955]. In a few cases measurements of the spectra have been made as a function of temperature, and these indicate the closing off at high temperatures of the region of transparency for glasses containing chromium or lead. *Clark* [1957b] measured the absorption spectra of olivine, diopside, pyrope, almandine, and grossularite in the visible and near infrared. The spectra of olivine and diopside are similar. There is a peak at about a micron due to the ferrous ion. The ultraviolet absorption edge lies at about 0.35 microns. All the minerals showed regions of transparency in which the absorption coefficients were a few cm^{-1}. Clark concluded that ferromagnesium silicates are sufficiently transparent for radiation to make an important contribution to the thermal conductivity. The results obtained by Clark suggest a value of about 10 cm^{-1} or less as an appropriate one for the opacity of silicates at room temperatures. Using measured values for the absorption coefficient, Clark calculated the radiative conductivity for the materials studied as a function of temperature. These calculations are listed in Table 10. It should be noted that for all materials the radiative conductivity is equal to or greater than the ordinary conductivity at temperatures of the order of 1500 to 2000°K.

TABLE 10—*Radiative conductivity calculated from absorption spectra [Clark, 1957b]*

	K_R, joules/cm sec °C			
Mineral	at 1000°K	at 1500°K	at 2000°K	at 2500°K
Olivine	0.297	0.86	1.45	2.02
Diopside	0.067	0.238	0.444	0.725
Pyrope	0.004	0.021	0.075	0.184
Almandine	0.004	0.017	0.042	0.212
Grossularite	0.047	0.193	0.456	0.405

In semiconductors the change of the absorption edge with temperature is on the order of -4 to -8×10^{-4} ev/°C. *Balchan and Drickamer* [1959] demonstrated a linear red shift of -4.2×10^{-4} ev/°C in olivine for temperatures up to 327°C. The change of the energy gap with pressure can be either positive or negative and is on the order of a few times 10^{-6} ev/bar. *Suchan and others* [1959] found that pressure decreases the energy gap in selenium, arsenic, white phosphorus, and iodine. A pressure of 1.5×10^5 bars shifts the absorption edge of olivine by about 9 per cent [*Balchan and Drickamer,* 1959]. It does seem probable that both temperature and pressure will tend to move the absorption edge towards longer wavelengths and to reduce the region of transparency. The values quoted above, however, are still consistent with the hypothesis that there is a gap between the infrared absorption and the ultraviolet absorption even in the deep mantle, though the absorption edge may lie in the visible. The effect of pressure and temperature on the infrared absorption is less than on the ultraviolet absorption edge [*Fan,* 1955].

Semiconductors show a marked increase of electrical conductivity with temperature. If this increase is interpreted in terms of intrinsic conduction by free electrons, then the opacity will markedly increase with an increase in temperature. Equation (6) illustrates the exponential dependence of the opacity on temperature through the mechanism of electrical conductivity.

In the calculations on the thermal history we shall assume that the best estimate of the opacity is 10 cm^{-1}, though we shall investigate the effect of higher opacities on the thermal history. The value of 10 cm^{-1} is considerably higher than that obtained by Clark for olivine. Most effects discussed in the preceding section tend to increase the opacity of silicates. The only effect that is explicitly allowed for in the calculations is the increase in general absorption due to free electrons.

In estimating the electrical conductivity we shall assume that our energy gap E is constant throughout the mantle. This is hardly to be expected, since both temperature and pressure will tend to decrease E, but in a manner as yet uncertain. Changes in phase and composition will further alter E. We shall usually assume that σ_0 is constant through the mantle. Both a decreasing energy gap and an increasing σ_0 in the mantle lead to an increased electrical conductivity. But these quantities have different effects on the radiative conductivity. A decrease in the energy gap will lead to an increase in the absorption. An increase in σ_0 implies a high mobility of the electrons rather than more electrons in the conduction band. The effect of a high value of σ_0 is to decrease the radiative conductivity, but to a lesser extent than does a change in the energy gap.

Electrical conductivity of the mantle—A knowledge of the electrical conductivity of the materials that make up the earth's mantle is important in discussions of the thermal state of the earth for two reasons: (1) The conductivity may give information as to the distribution of temperature, provided that the mechanism of electrical conduction is known. (2) The electrical conductivity enters into the expression for the variation of thermal conductivity with temperature.

The electrical conductivity of the earth's mantle can be inferred from a study of geomagnetic transient variations. The periodic solar daily variations and the aperiodic magnetic storms can be separated by a Gaussian analysis into induced and exciting components. A knowledge of the total field at the earth's surface enables one to compute the maximum depth penetrated by the induced currents. Knowing the depth of penetration of the induced currents over a wide range of frequency, one can calculate the distribution of the conductivity. *Lahiri and Price* [1939], using the short-period variations in the magnetic field, estimated the conductivity to a depth of about 800 km. They showed that a large number of interpretations are compatible with the magnetic variations, but they also showed that there is a very rapid increase in conductivity somewhere within the outer few hundred kilometers of the earth. One distribution of conductivity obtained by Lahiri and Price is given in Table 11; it corresponds to their curve *d*. An alternative distribution is one in which the radial variation of the conductivity becomes infinite at a depth of about 600 km. The principal difficulty in interpreting the short-period variations in the magnetic field of external origin results from the uncertain shielding effect of the oceans. The analysis by *Rikitake* [1951] on the effect of a sea bounded by two meridians supports the conclusion of Lahiri and Price that the conductivity rises sharply somewhere in the outer few hundred kilometers of the earth.

TABLE 11—*Electrical conductivity of mantle*

Depth, km	σ_0,* ohm^{-1} cm ^{1}	Range of uncertainty of σ, ohm^{-1} cm^{-1}	σ,† ohm^{-1} cm^{-1}
200	5×10^{-6}		1×10^{-4}
400	8×10^{-5}		3×10^{-4}
600	7×10^{-4}		0.0016
900	0.04		0.008
1400	0.35	0.14-0.76	
1900	0.60	0.27-1.1	
2400	1.0	0.4-2.8	
2800	2.2	0.6-7.0	

**McDonald* [1957].

†*Lahiri and Price* [1939], curve *d*.

McDonald [1957] extended the analysis of Lahiri and Price to longer-period variations of internal origin of the magnetic field and was thus able to estimate the conductivity at the core-mantle boundary. His values are shown in Table 11 together with his estimated range of uncertainty. McDonald found that the electrical conductivity increases gradually with depth in the lower mantle to a value on the order of 2 ohm^{-1} cm^{-1} at the core-mantle boundary. McDonald also estimated the conductivity near the surface by means of the short-period variations of external origin. His values lie within the range indicated by Lahiri and Price.

Coster's [1948] laboratory studies established the strong temperature dependence of electrical conductivity in silicates. *Hughes* [1953] confirmed Coster's results and further showed that the temperature dependence of conductivity can be interpreted in terms of three different mechanisms: impurity conduction, intrinsic semiconduction, and ionic conduction. The first two mechanisms obey the relation

$$\sigma_0 e^{-E/2kT} \qquad (7)$$

where E is the energy gap of the crystal and σ_0 is a constant of proportionality with the interpretation of conductivity at infinite temperature. The third mechanism obeys a similar relation but without the factor of ½ in the exponential. The numerical values of σ_0 and E as obtained by Hughes for impurity, intrinsic, and ionic conductivity of olivine are listed in Table 12 together with the range of temperature where the given mechanism is important. Hughes found similar results for other ferromagnesium minerals.

There is disagreement as to the mechanism which causes the conductivity of the mantle. *Hughes* [1959] suggested that it is due to ionic conductivity. Runcorn, Tozer, and Clark have all interpreted the conductivity in terms of intrinsic semiconduction. Determination of the mechanism of conductivity is important in attempting to estimate the temperature distribution from estimates of conductivity, as Tozer has done. For the problem of thermal conduction it is unimportant which mechanism of conductivity is *dominant*. The decrease in heat transported by radiation is due primarily to free electrons and is thus dependent upon intrinsic semiconduction. The contribution of ionic conductivity to electrical conductivity does not appreciably lower the amount of energy that can be transported by radiation. In calculations of the thermal history we have assumed that the semiconductivity of the material in the mantle can be approximated if we use Hughes' values for intrinsic semiconduction in olivine.

TABLE 12—*Electrical conductivity of olivine*

Type of conduction	σ_0, ohm^{-1} cm^{-1}	E, ev	Range of dominance, °C
Impurity	10^{-4}	1	<600
Intrinsic	10	3.2	600-1100
Ionic	10^6	3.0	>1100

Melting relations at high pressures—Since the mantle of the earth is solid, knowledge of the melting temperature of the material of the mantle would effectively limit estimates of possible temperature distributions within the outer part of the earth. At present, the pressure variation of the melting temperature of geologically important materials is one of the most uncertain of all geophysical parameters. Recent high-pressure experiments at the General Electric Company and the Geophysical Laboratory of the Carnegie Institution of Washington provided important data which permit an estimate of the initial slope of the melting-point curves. Extrapolations of these initial slopes to conditions existing within the deeper earth require some sort of theory. It is customary to use Simon's semi-empirical equation, which has proved valuable in summarizing melting-point data of inert gases.

Yoder [1952] found that the melting point of diopside increases with pressure at the rate of 13°C per 1000 bars for pressures up to 5000 bars. *Boyd and England* [1958] extended Yoder's work to 30,000 bars. In the pressure range of 20,000 to 30,000 bars the average slope is 10.3°C per 1000 bars. This work permits an estimate not only of the initial slope of the melting curve for diopside but also of the initial change of slope. P. LeComte (personal communication) obtained an initial slope of 11° ± 2°C. per 1000 bars for the albite melting curve. *Strong* [1959] determined that a pressure of 96,000 atmospheres raises the melting point of iron by 190° ± 20°C.

The Simon semi-empirical equation relates the pressure and temperature along the fusion curve by

$$P = (a/B)[(T/T_0)^B - 1] \qquad (8)$$

where a and B are empirical constants. T_0 is the melting temperature at 1-bar pressure. The initial slope of the melting-point curve can be used to determine the constant a, and B is determined by the initial curvature. *Gilvarry* [1956a,b,c,d] examined the relation of the Simon equation to the Debye theory of simple solids. He found that the Simon equation is consistent with the Debye theory, provided that Poisson's ratio of the solid is constant or slowly varying along the fusion curve. The theory is developed for a monatomic solid in which the Debye temperature is taken as a constant. Gilvarry has shown that for monatomic solids the exponent B in the Simon equation can be related to Grüneisen's constant γ by $B = (6\gamma + 1)/(6\gamma + 2)$. Polyatomic materials, such as silicates, cannot be treated in terms of such a simple theory, and the 'Debye temperature' itself is a function of both pressure and temperature [*MacDonald*, 1956]. Since the theoretical foundations of the Simon equation as applied to silicates are very much in doubt, and since there is no experimental justification for its applicability to silicates, extrapolations based on the Simon equation must be viewed with reserve.

The coefficients for the Simon equation for various materials are listed in Table 13. In the case of diopside, the constants a and B were taken from the initial slope and the initial curvature of the melting-point curve. The initial slope of albite determines the constant a. The constant B was obtained from Grüneisen's parameter evaluated at the melting temperature. Two sets of constants for iron are listed. One set was determined by Strong from his experimental data. The derived Grüneisen's parameter is much smaller than that obtained by other methods. The second set was determined by means of Grüneisen's parameter for iron derived from shock-wave experiments [*Altshuler and others*, 1958] and from the initial slope of the fusion curve. These parameters also fit Strong's data to within the range of experimental error. Strong's extrapolation leads to a melting point of iron at the core-mantle boundary of 2340° ± 200°C. whereas the second set of parameters gives 3920°C. The two divergent estimates illustrate the difficulties of extrapolation when the curvature of the melting-point curve is small (Fig. 1).

TABLE 13—*Coefficients for Simon's equation*

Mineral	$10^{-5}\,a$ bars	B	T_0, °K	γ at T_0
Diopside	1.28	2.6	1664	0.7
Albite	1.26	2.8	1391	0.6
Iron*	0.75	8	1805	0.24
Iron†	9.0	1.4	1805	1.6

*From *Strong's* [1959] melting point data.

†B determined from Grüneisen's parameter obtained from shock-wave measurements; a fixed by slope of melting-point curve.

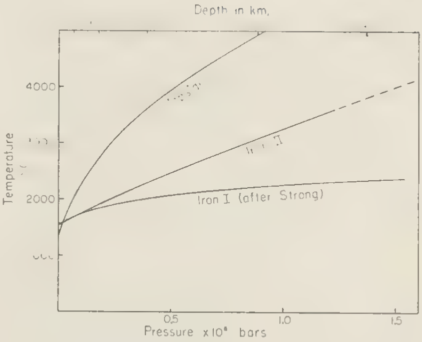


FIG. 1—Extrapolated melting relations for iron and diopside within the earth. Iron I curve is *Strong's* [1959] estimate. Iron II curve is extrapolated by using the initial slope obtained by Strong and Gruneisen's constant as appropriate at high pressures.

The melting points of minerals as estimated from Simon's equation at various depths below the earth's surface are listed in Table 14. Figure 1 illustrates the possible extrapolation of Strong's melting-point data together with the projected diopside fusion curve. Because of the extreme uncertainties involved in the extrapolation of the initial values of the slope and

Table 14—*Melting temperatures of minerals estimated from Simon's equation**

Pressure, megabars	Equivalent depth in earth, km	Diopside m.p., °C	Albite m.p., °C	Iron† m.p., °C	Iron‡ m.p., °C
0.1	300	2250	1825	1720	1750
0.21	600	2725	2300	1860	1960
0.39	1000	3600	2800	2000	2280
0.63	1500	4300	3450	2140	2680
0.88	2000	4925	3800	2220	3100
1.37	2900	...	...	2340	3920

*It is assumed that if the minerals undergo a phase change at high pressure, the phase boundary does not intersect the melting curve. Albite is certainly not stable at depths greater than about 50 km, and diopside may also undergo a phase change.

†*Strong* [1959].

‡Derived from second set of parameters for iron in Table 13.

curvature of the melting curves, the temperatures listed are no more than an indication of the temperatures to be expected. Yet it is interesting to note that iron has by far the lowest melting temperature, and the melting temperature of albite lies well below that of diopside throughout the earth's mantle.

In a sense, the melting points listed in Table 14 are estimates of the maximum temperatures that could exist within the solid mantle. Furthermore, the mantle is undoubtedly a multicomponent system. Melting relations in such systems are not simple, but it is certain that melting would begin at a temperature lower than the melting temperature of any single phase in the system, further lowering the maximum possible temperature.

Part II: Calculations and Descriptions of Models

Difference equation and numerical stability—In treating the thermal history we assume an initial temperature distribution, a distribution of heat sources, and a distribution of the parameters determining density, heat capacity, and thermal conductivity. The temperature at the surface of the earth is assumed to be known and constant in time. The problem is then to determine the distribution of temperature at later times. If volume changes associated with changes in temperature are neglected, the equation governing the temperature distribution in a spherically symmetric body is

$$\rho C_p \frac{\partial T}{\partial t}(r, t) = \frac{1}{r^2}\frac{\partial}{\partial r} \cdot\left\{r^2 K \frac{\partial T}{\partial r}(r, t)\right\} + A(r, t) \qquad (9)$$

The rate of heat production per unit volume A is a function of time as well as of radius, since, if a radioactive element is producing heat at a rate dQ/dt today, it produced heat at a rate $e^{\lambda t}$ (dQ/dt) at a time t years ago, where λ is the disintegration constant of the radioactive element. Analytical solutions to this equation can be obtained for special problems if the thermal conductivity is assumed to be independent of temperature [*Carslaw and Jeager*, 1959]. If radiative transfer of heat is important, the conductivity depends on temperature and the equation becomes a nonlinear partial differential equation. In this case there are no known solutions for the general problem, and numerical methods are indicated.

Let Δr and Δt be increments of the variables r and t. The set of points in the r-t plane given by

$$\begin{aligned} r &= m\Delta r \quad m = 0, 1, \ldots, M; \; M\Delta r = R \\ t &= n\Delta t \quad n = 0, 1, \ldots \end{aligned} \qquad (10)$$

determines a grid whose mesh size is fixed by Δt and Δr. R is the radius of the spherical body. Let

$$\begin{aligned} T_m^n &= T(n\Delta t, m\Delta r), \quad K_m = K(m\Delta r) \\ A_m^n &= A(n\Delta t, m\Delta r), \quad \rho_m = \rho(m\Delta r) \end{aligned} \qquad (11)$$

A derivative in time is approximated by

$$(T_m^{n+1} - T_m^n)/\Delta t \qquad (12)$$

and a second derivative with respect to radius is approximated by

$$(T_{m+1}^n - 2T_m^n + T_{m-1}^n)/(\Delta r)^2 \qquad (13)$$

These are the simplest finite difference approximations. More complicated and accurate representations are possible. The finite difference equation approximating equation (9) is then

$$\begin{aligned} T_m^{n+1} = T_m^n + \frac{\Delta t}{\rho_m C_p} \Bigg\{ & \frac{1}{m(\Delta r)^2} \\ & \cdot \left(\left[\frac{m}{4} (K_{m+1} - K_{m-1}) \right. \right. \\ & \left. + (m + 1) K_m \right] T_{m+1}^n - 2mK_m T_m^n \\ & - \left[\frac{m}{4} (K_{m+1} - K_{m-1}) \right. \\ & \left. \left. - (m - 1) K_m \right] T_{m-1}^n \right) + A_m \Bigg\} \end{aligned}$$

$$\text{for } m \neq 0 \qquad (14a)$$

$$\begin{aligned} T_0^{n+1} = T_0^n + \frac{\Delta t}{\rho_0 C_p (\Delta r)^2} & [6K_0(T_1^n - T_0^n) \\ & + (K_1 - K_0)(T_1 - T_0)] + \frac{A_0 \Delta t}{\rho_0 C_p} \end{aligned}$$

$$\text{for } m = 0 \qquad (14b)$$

with the external condition of

$$T_m^n = 0 \qquad (15)$$

and the initial condition of

$$T_m^0 = f(m\Delta r) \qquad (16)$$

where f is a given function. By using the forward time differences we employ the current values of the temperature at given space points to predict ahead in time.

It would appear that the approximation of the finite difference equation to the partial differential equation could be improved to any desired degree by making the time and space intervals smaller and smaller. This is not true for simple diffusion equations, as was proved by *Courant and others* [1928]. Errors in the approximation can be made worse by taking smaller values of the space intervals, unless the time interval is also suitably reduced. For the ordinary diffusion equation

$$\partial T/\partial t = \sigma(\partial^2 T/\partial x^2) \qquad (17)$$

the condition for the stability of the difference equation approximation is

$$(\sigma \Delta t)/(\Delta x)^2 \leq 1 \qquad (18)$$

Richtmyer [1957] presented examples in which this stability condition is not met and in which case the approximate numerical solution shows undamped oscillations about the true solution. He further showed that lower-order terms in the diffusion equation do not affect the stability.

The condition for the stability of the approximation given in (18) is based on a constant diffusivity σ. For the case of variable conductivity and density an empirical approach has been used. Several numerical experiments were run in which the difference equation (14) was used and in which it was assumed that the conductivity was proportional to the cube of the temperature. It was found that as long as the condition in (18) was met, the finite difference approximation did not show undamped oscillations, but if this condition was not satisfied the difference equation showed instability in the form of large-amplitude oscillations about the solution.

In the actual computations on the thermal history, the program was designed to test for the condition

$$\frac{K_m}{\rho_m C_p} \frac{\Delta t}{(\Delta x)^2} \leq \tfrac{1}{2} \qquad (19)$$

at each iteration. If this condition was not met, the program automatically adjusted the space and time differences to meet the stability requirement. Since a major interest in the problem is the development of the temperature distribution in the outer few hundred kilometers, the space steps were initially taken as

$$\Delta x = 100 \ km \qquad (20)$$

For the highest conductivities reached in the problem, the time steps then satisfied the condition

$$\Delta t \leq 1.4 \times 10^7 \text{ years} \qquad (21)$$

A refinement of the space net to intervals of less than 100 km would have required much smaller time intervals in order to satisfy equation (19). This would have vastly increased the time needed for the computation. There are no theoretical limits to the magnitude of the space interval for steady-state problems.

The results of a large number of computations are presented in numerical and graphical form. It should be noted that the calculations have been carried out for *assumed* models of the earth. The emphasis is on the *general* features of *possible* thermal histories of the earth rather than on the detailed numerical estimates. These estimates are no better than the parameters and approximations used in the calculations.

Near-surface thermal conditions—The distribution of temperature near the outer surface of the earth depends primarily on the concentration of radioactive materials in the outer layers [*Slichter,* 1941; *Bullard,* 1954; *Birch,* 1955]. The initial thermal state of the earth has little influence on the distribution of temperature in the outer few tens of kilometers because of the short thermal time constant for this region. A first approximation of the temperature distribution in the outer layers of the earth can be obtained by examining the steady-state temperature distribution.

Since we are considering the distribution of temperature over a distance that is small compared with the radius of the earth, it is sufficient to consider the distribution of temperature in a slab in which heat is being produced by radioactive decay. If A is the instantaneous rate of heat production per unit volume and per unit time, the instantaneous equilibrium distribution of temperature is determined by

$$K(d^2T/dz^2) = -A(z, t) \qquad (22)$$

Since the rate of heat production varies with time, we investigate a quasi-stationary distribution of temperature. The time constant associated with heat production is long compared with the thermal time constant of a slab 30 km thick. The solution to the differential equation is

$$T = \tfrac{1}{2}(Az^2/K) + Bz + C \qquad (23)$$

where z is the depth measured from the surface. The constant B is determined by the surface heat flow and for the earth is equal to

$$B = \frac{\text{Heat flow}}{K} \sim \frac{50}{K} \qquad (24)$$

C is the temperature at the surface. From (23) we see that the temperature at a given depth z is a maximum if there is no heat production in the material between the surface and depth z. The temperature at the depth z is minimum if all production is concentrated in the layer between the surface and the depth z. This minimum temperature is one half of the maximum temperature. For a heat flow of 50 ergs/cm^2 sec the temperature at a depth of 30 km is 620°C, on the assumption that the surface temperature is 20°C, that the conductivity is 0.025 joules/cm sec °C, and that there is no heat produced over the 30-km interval. If all the heat that gives rise to the heat flow of 50 ergs/cm^2 sec is produced within the 30-km slab, then the temperature at 30 km is 320°C. These figures then indicate the approximate range of temperatures that might be expected at the base of a continental crust 30 km thick.

A further indication of the temperatures that might exist at the base of the crust can be found if we use simple models of the crust and the estimated conductivity and rates of heat production listed in Tables 9 and 7. A continental crust is assumed to be 30 km thick with a surface temperature of 20°C. If the crust is made up of granitic material, then the heat produced within the crust is equivalent to a heat flow of 60 ergs/cm^2 sec, which is somewhat larger than the observed heat flow. If the crust is made up of intermediate rock, the heat flow at 30 km is 40 per cent of the total heat flow, with the crustal materials producing 60 per cent of the heat. The temperature at a depth of 30 km is 345°C or 415°C depending upon the particular conductivity assumed (Table 15). If the continental crust is made up of material having a heat production equivalent to that of a basalt, then 60 per cent of the heat comes from a depth below 30 km. The temperatures now are 500°C or 600°C, again depending on the particular value for the thermal conductivity. At these low temperatures the conductivity is due to transfer of energy by lattice vibration, with the radiative component being negligible.

Table 15—*Near-surface thermal conditions for a surface heat flow of 50 ergs/cm² sec*

Region	Type of rock (see Table 5)	Heat produced above 30 km, ergs/cm³ year	Conductivity, joules/cm sec °C	T, °C at 30 km depth	Heat flow at 30 km, ergs/cm² sec
Continental	Intermediate	302	0.033	345	21
Continental	Intermediate	302	0.027	415	21
Continental	Basalt	209	0.025	500	30
Continental	Basalt	209	0.020	615	30
Oceanic	Basalt	209	0.025	430	33
Oceanic	Basalt	209	0.020	535	33
Oceanic	Hualalai basalt	107	0.025	475	41
Oceanic	Hualalai basalt	107	0.020	590	41

For the model oceanic region we assume 4 km of water with a temperature of 0°C at the base of the ocean. The heat productivity across the Mohorovicic discontinuity is assumed to be constant. A material having the heat productivity of the Hualalai basalt would give a temperature at a depth of 30 km of 475°C to 590°C, with 80 per cent of the total surface heat flow of 50 ergs/cm² sec coming from below a depth of 30 km. If the transition at the base of the oceanic crust is from basaltic material to dunite, then the temperatures reached at 30 km are considerably higher. The rate of heat production of dunite appears to be very much less than that of basalts, though this effect is partly compensated for by the higher conductivity of the dunites.

In both continental and oceanic regions the temperature at a depth of 30 km should be of the order of 500°C. The exact value depends both on the distribution of radioactivity in the crustal materials and on the distribution of conductivity. A temperature much in excess of 600 to 700°C appears unreasonable. Local variations are to be expected, since the influence of sediments of low conductivity and of a large horizontal extent will be felt at shallow depths. As an example, temperatures at a depth of 30 km in areas of Tertiary sedimentation should be a hundred degrees or more above temperatures in ancient shield areas.

Steady-state temperature distribution—Birch [1958] noted that the surface heat flow of 50 ergs/cm² sec was consistent with the hypothesis that the earth has a chondritic composition. If the coincidence in values between the present rates of heat production and heat flow is significant, then the steady-state distribution of temperatures within the earth should give a first-order estimate of the actual temperature distribution. There is an important alternative reason for study equilibrium temperature distributions. Such a study illustrates the dependence of the temperature distribution on the parameters that determine the total thermal conductivity. In a steady-state treatment the complications associated with the time-varying quantities are eliminated. The steady-state distribution also focuses attention on the importance of the distribution of heat sources.

For a spherically symmetrical earth the steady-state distribution of temperature (if the variations in the rate of heat production are neglected) is determined by

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 K \frac{\partial T}{\partial r}\right) = -A(r) \qquad (25)$$

where A is the rate of heat production per unit volume and K is the thermal conductivity. The present rate of heat production in chondrites is about one third of the average rate of heat production over the past 4.5×10^9 years. Using present day values to determine A implies a neglect of the heat produced by 'fossil radioactivity.' The thermal conductivity is assumed to be of the form

$$K = C + \frac{16n^3 sT^3}{n\epsilon_0 + 120\pi\sigma_0 e^{-E/kT}} \qquad (26)$$

where ϵ_0 is the opacity at low temperature, E is the width of the energy gap for electronic conduction, s is the Stefan-Boltzman constant, n is the index of refraction, and k is Boltzman's constant. C is the ordinary conductivity at low temperatures. It is taken to be a constant equal to 0.025 joules/cm sec °C. The index of refraction is also taken as a constant equal to 1.7. The electronic energy gap is also assumed constant throughout the mantle and equal to 3 ev. The present rate of heat production in the model earth (Table 4) is used in determining the heat source distribution $A(r)$.

The time-independent part of equation (16) has been programed for an IBM 704, and the equilibrium temperature distributions were obtained by using a space interval of 1 km. The effects of the variation of radioactivity, opacity, and σ_0 are presented below.

Table 16 illustrates the dependence of the equilibrium temperature distribution on the value of the zero temperature opacity and on the nature of the mechanism for electrical conduction within the mantle. In the three models listed in Table 16 the radioactivity is assumed to be uniformly distributed and to be such that the heat flow at 30 km is equal to 39.5 ergs/cm² sec. In model 1 the opacity is taken at 30 cm⁻¹. The temperature at the core-mantle boundary is in excess of 7000°C. (The core is assumed to be isothermal.) Comparing the temperatures listed with estimated melting-point temperatures (Fig. 1), one sees that such a distribution of conductivity and radioactivity would indicate a mantle largely or completely molten. In model 2 (Fig 2) the opacity is 10 cm^{-1}. This reduction by a third in the opacity reduces the temperature of the core boundary by some 1400°C. The thermal conductivity reaches a limiting value in which the increase of thermal conductivity due to the increase in temperature is balanced by the decreasing conductivity due to the increased electrical conduction. In both of these cases electrical conductivity has the striking effect of reducing the rate of increase of the total conductivity at depths of 700 to 1000 km.

TABLE 16—*Steady-state temperatures in a mantle having uniform heat production*

Depth, km	Model 1 T, °C	Model 2 T, °C	Model 3 T, °C
120	1520	1300	1300
320	2700	2110	2090
520	3420	2600	2510
720	4000	3020	2810
920	4510	3400	3040
1120	4970	3770	3240
1320	5380	4110	3420
1520	5750	4440	3570
1720	6070	4730	3701
1920	6340	4990	3810
2120	6570	5210	3900
2320	6750	5390	3980
2520	6890	5530	4030
2720	6970	5610	4060

Description of models

1. T at 30 km = 420°C. Heat flow at 30 km = 39.5 ergs/cm² sec. ϵ_0 = 30 cm^{-1}, σ_0 = 10 ohm^{-1} cm^{-1} throughout mantle.
2. T at 30 km = 420°C. Heat flow at 30 km. = 39.5 ergs/cm² sec. ϵ_0 = 10 cm^{-1}, σ_0 = 10 ohm^{-1} cm^{-1} throughout mantle.
3. T at 30 km = 420°C. Heat flow at 30 km = 39.5 ergs/cm² sec. ϵ_0 = 10 cm^{-1}, throughout mantle. σ_0 = 1 ohm^{-1} cm^{-1} from 0 to 1000 km depth, σ_0 = 2 ohm^{-1} cm^{-1} from 1800 to 2889 km.

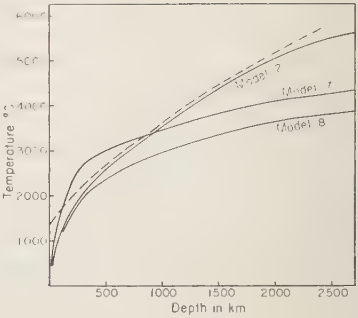


FIG. 2—Steady-state temperature distributions. Extrapolated diopside melting curve shown by dashed line. Model 2 represents an undifferentiated mantle having a low opacity (Table 17). Models 7 and 8 represent a partially differentiated mantle (Table 18).

Model 3 differs from models 1 and 2 in that the electrical conductivity at depth is assumed to be lower by taking a lower value of σ_0. In this case the thermal conductivity continues to increase throughout the mantle and the highest temperature is 1600°C lower than that attained for a model having an identical opacity but having a higher value of electrical conductivity. The highest thermal conductivity of 0.45 joules/cm sec °C is almost 20 times as great as the ordinary lattice conductivity.

All three models are similar in that the temperature gradient is greatest near the surface and becomes less as the depth increases. This behavior is to be expected from the strong temperature-dependence of the conductivity. At higher temperatures the flow of heat is increased by the higher conductivity, and the temperature gradient required to remove the heat becomes proportionally less.

None of the models considered in Table 16 can be thought of as appropriate to the earth. For the highest conductivities the time constant for a spherical shell 2800 km thick is 30 × 10^9 years, far in excess of the age of the earth. If thermal conduction is the only mechanism of heat transport, it is impossible for the material within the deep mantle to have come to thermal equilibrium with the surface.

Three cases in which the radioactivity is concentrated within the upper 430 km of the earth are listed in Table 17. Model 4 approximates an oceanic case in which the heat flow at 30 km is 54 ergs/cm^2 sec and the opacity is 30 cm^{-1}. The temperature reached at a depth of 420 km is 2700°C. In model 5 we approximate a continental region with a heat flow at a depth of 30 km of 26 ergs/cm^2 sec. In this case the higher crustal concentration of radioactivity leads to a lower temperature than in the case of model 4. The temperature at 420 km is 1927°C. Model 6 is similar to model 4, except that the opacity is taken to be one-third that of model 4. There is a resulting decrease of some 600°C in the equilibrium temperature at 420 km.

TABLE 17—*Steady-state temperatures in mantle in which radioactivity is in the upper 430 km*

Depth, km	Model 4 T, °C	Model 5 T, °C	Model 6 T, °C
70	1140	820	1020
120	1750	1180	1430
170	2070	1440	1680
220	2320	1620	1840
270	2490	1760	1960
320	2610	1850	2030
370	2680	1900	2080
420	2710	1930	3000

Description of models

4. T at 30 km = 420°C. Heat flow at 30 km = 54 ergs/cm^2 sec. ϵ_0 = 30 cm^{-1}, σ_0 = 10 ohm^{-1} cm^{-1}.
5. T at 30 km = 460°C. Heat flow at 30 km = 26.1 ergs/cm^2 sec. ϵ_0 = 30 cm^{-1}, σ_0 = 10 ohm^{-1} cm^{-1}.
6. T at 30 km = 420°C. Heat flow at 30 km = 54 ergs/cm^2 sec. ϵ_0 = 10 cm^{-1}, σ_0 = 1 ohm^{-1} cm^{-1}.

Models 5 and 6 represent equilibrium distributions which satisfy the surface heat flow and the requirements of a solid mantle yet have equilibrium temperature distributions. An additional rise in temperature due to adiabatic compression might be sufficient to raise the temperature at the core-mantle boundary above the melting point of iron. The principal difficulties in these two models are the low electrical conductivity that would be reached at the core-mantle boundary and the small rate of increase of conductivity in the region from 400 to 1000 km. Both of these requirements could be met if additional variation in the parameters affecting the electrical conductivity were allowed. Furthermore the effects of initial heat and fossil radioactivity have been neglected.

Temperatures for a partially differentiated mantle are given in Table 18. In all these models the uranium and thorium are assumed to be concentrated above 400 km and there is a partial concentration of potassium above 400 km. It is assumed that the concentration of potassium in the material below 400 km is 4.2 × 10^{-4} g/g compared with the concentration in chondritic material of 8.0 × 10^{-4} g/g. This concentration of potassium gives rise to a present-day heat production of 1.5 × 10^{-8} ergs/g sec. The heat flow at 400 km due to the deeply buried potassium is 11.2 ergs/cm^2 sec. The requirements of a surface heat flow of about 50 ergs/cm^2 sec are satisfied by these models, and since the major proportion of the heat is concentrated above 400 km the models are consistent with a rough equality of chondritic heat production and surface heat flow. The resulting temperatures are higher than those obtained in models 4, 5, and 6 due to the deeply buried potassium. The actual values reached depend upon the particular parameters assumed in the expression for the radiative conductivity. Constant values of opacity and σ_0 throughout the mantle are assumed in models 7 and 8 (Fig. 2). Model 9 has a constant opacity and a low value for σ_0. Model 10 has a varying opacity. The temperatures obtained in models 8 and 10 are somewhat less than the estimated melting temperatures, though the uncertainty of the latter prevents any

Table 18—*Steady-state temperatures for partially differentiated mantle*

Depth, km	Model 7 T, °C	Model 8 T, °C	Model 9 T, °C	Model 10 T, °C
120	1730	1200	1730	1440
320	2730	2030	2690	2110
520	3030	2390	2950	2360
720	3230	2650	3110	2610
920	3410	2870	3250	2800
1120	3570	3060	3380	3050
1320	3710	3230	3490	3340
1520	3850	3370	3590	3580
1720	3960	3500	3670	3920
1920	4060	3610	3740	4050
2120	4150	3700	3800	4150
2320	4220	3780	3850	4220
2520	4270	3830	3880	4260
2720	4300	3860	3900	

Description of models

7. T at 30 km = 420°C. Heat flow at 30 km = 54 ergs/cm² sec, at 400 km = 11.2 ergs/cm² sec. Radioactive heat production above 400 km = 3.4×10^{-7} ergs/g sec, from 2890 to 400 km = 1.5×10^{-8} ergs/g sec. $\epsilon_0 = 30$ cm^{-1}, $\sigma_0 = 10$ ohm^{-1} cm^{-1} throughout mantle.
8. T at 30 km = 460°C. Heat flow at 30 km = 26 ergs/cm² sec, at 400 km = 11.2 ergs/cm² sec. Radioactive heat production above 400 km = 1.2×10^{-7} ergs/g sec, from 2890 to 400 km = 1.5×10^{-8} ergs/g sec. $\epsilon_0 = 30$ cm^{-1}, $\sigma_0 = 10$ ohm^{-1} cm^{-1} throughout mantle.
9. T at 30 km = 420°C. Heat flow at 30 km = 54 ergs/cm² sec, at 400 km = 11.2 ergs/cm² sec. Radioactive heat production above 400 km = 3.4×10^{-7} ergs/g sec, from 2890 to 400 km = 1.5×10^{-8} ergs/g sec. $\epsilon_0 = 30$ cm^{-1} throughout mantle. $\sigma_0 = 1$ ohm^{-1} cm^{-1} above 1000 km, $\sigma_0 = 2$ ohm^{-1} cm^{-1} from 2890 to 1000 km.
10. T at 30 km = 420°C. Heat flow at 30 km = 54 ergs/cm² sec, at 400 km = 11.2 ergs/cm² sec. Radioactive heat production above 400 km = 3.4×10^{-7} ergs/g sec, from 2890 to 400 km = 1.5×10^{-8} ergs/g sec. $\epsilon_0 = 10$ cm^{-1} above 400 km, $\epsilon_0 = 30$ cm^{-1} from 400 – 1000 km, $\epsilon_0 = 100$ cm^{-1} from 1000 to 2890 km, $\sigma_0 = 1$ ohm^{-1} cm^{-1} above 1000 km, $\sigma_0 = 2$ ohm^{-1} cm^{-1} from 2890 to 1000 km.

definite statement. The electrical conductivities deep within the earth which correspond to the models are low compared with the inferred conductivities. Again, this difficulty can be resolved by assuming higher values for σ_0 or lower values for E. Neither of these alterations would seriously affect the final temperature, since the principal variation in temperature is within the upper part of the mantle. In all these models the lower part of the mantle would not have reached thermal equilibrium during the lifetime of the earth. However, in the case of models 4 to 10 the concentration of radioactivity towards the surface insures that most of the heat being produced within the earth is reaching the surface at an equilibrium rate.

A necessary requirement for a present steady-state approximation is that a major portion of any initial heat must have been lost by processes other than ordinary thermal conduction and radiative transport. Furthermore, according to the chondritic hypothesis for the composition of the earth, the present rate of heat production is a factor of 3 lower than the average rate over the last 4.5×10^9 years. The present coincidence of heat production and heat loss requires that the heat produced at a higher rate in the past be lost through near-surface concentration of radioactivity or by convective transport of heat. Various distributions of radioactivity and thermal parameters are in agreement with the hypothesis of a thermal steady state for the earth. But these suggestions are unrealistic until the effects of initial heat and fossil radioactivity are critically examined.

Initial conditions—A solution to the initial-value problem for the thermal history of the earth requires specification of the distribution of temperature within the earth at zero time. This specification requires an assumption for that part of the earth's history about which we know the least. Indeed it is one of the aims of the present study to obtain limits on the possible conditions that may have existed during the time the earth formed. For the sake of definiteness in our arguments we shall assume that the earth mechanically aggregated by some process of accretion. The accretion hypothesis as put forward by Spitzer, Hoyle, Urey, and others has gained a wide degree of acceptance in spite of the many problems still unsolved.

For our purposes the important parameter is the length of time in which the earth accreted. Most of the accepted values for the time of accretion center about 10^8 years. That this time is reasonable can be seen from the following considerations. Consider a gas cloud at a uniform temperature. The inward velocity of any material particle at a distance r from the center of aggregation is

$$(2GM/r)^{1/2} \tag{27}$$

where M is the total mass of the cloud. The characteristic time associated with gravitational aggregation is then

$$t \approx r^{3/2}/(2GM)^{1/2} \tag{28}$$

Initial dimensions of the dust cloud are fixed if we suppose that densities observed in modern dust clouds are representative of the conditions that existed at the time of the formation of the earth. *Dufay* [1957] estimated densities on the order of 10^{-21} g/cc. The characteristic time for gravitational aggregation is then on the order of 2×10^7 years. *Hoyle* [1946] considered the condensation from a gaseous disk of material. For an object of mass smaller than 10^{23} grams condensation is important and the time scale is on the order of 10^8 years. For larger masses accretion dominates during the latter stages of aggregation and again the time scale is on the order of 10^8 years.

Possible heat sources that could raise the temperature of the earth during accretion include radioactive decay of both long-lived and short-lived isotopes, chemical reaction, conversion of kinetic energy into thermal energy, and release of energy stored by radiation damage in the aggregating material.

Suppose that the earth aggregated 4.5×10^9 years ago with the aggregation process taking place over a period of 10^8 years. We assume that the material which made up the earth had the composition of chondrites. The rate of energy released by long-lived radioactive isotopes during this time was on the average 12.8 ergs/g year (Table 4). If all the radioactively produced heat was retained by the aggregating particles, the increase in temperature in 10^8 years due to radioactivity of long-lived isotopes was then

$$\begin{aligned} T &\approx (dQ/dt)\Delta t/c_p \\ &\approx (12.8 \times 10^8)/(8.0 \times 10^6)\Delta T \\ &\approx 160°C \end{aligned} \tag{29}$$

where c_p is the specific heat.

Short-lived radioactive isotopes might have contributed to the initial heat of the earth if the time between the formation of the elements and the aggregation of the earth was short compared with the half-lives of the isotopes. An indication of the order of magnitude of the heat that could be generated by short-lived isotopes is provided if we assume that the time between element formation and aggregation was 10^8 years. The value of 10^8 years is suggested by *Wasserburg and Hayden* [1955b] since the abundance of xenon-129 in the Beardsley chondritic was below the limit of detectability (1.3×10^{-11}C c/g). We need to know the initial abundance of the isotopes at the time of formation. *Kohman* [1956] reported on a search for extinct radioactivity and *Burbidge and others* [1958] calculated abundances for the isotopes produced in a rapid explosive process. The isotopes having a half life of 10^6 years and greater are listed in Table 19, together with the disintegration energies and estimated abundances. The abundances of Be^{10} and Al^{26} have been estimated from their position on the curve of abundance versus mass number. The other estimates were made by Burbidge and his co-workers except for K^{40}, for which we use the chondritic abundance extrapolated back 4.5×10^9 years. The relative heat production of the various isotopes as compared with potassium-40 is also shown in Table 19. Al^{26} could have been a large source of energy during the first few hundred thousand years after the formation of the elements [*Kohman*, 1956]. However, because of the slowness of the

TABLE 19—*Relative heat production by short-lived isotopes during initial 10^8 years*

Isotope	Half-life, 10^6 years	Disintegration energy in Mev	Abundance 4.5×10^9 years ago (relative to Si-10^6)	Heat production relative to K^{40}
Be^{10}	2.7	0.56	(4.5)*	7.5
Al^{26}	~1	4.0	(1×10^3)*	1×10^3
Zr^{93}	~0.9	0.03	0.2	0.014
Pd^{107}	7	0.02	0.07	0.004
I^{129}	17	0.19	1.0	0.7
Ca^{135}	2.0	0.1	0.2	0.07
Sm^{146}	50	2.6	0.25	1.4
V^{236}	23.9	4.6	0.2	2.6
Po^{244}	75	15.2	0.4	11.0
Cm^{247}	40	16.3	0.1	4.0
K^{40}	1300	0.71	8.4	1.0

*Estimated from position on the curve for abundance versus mass-number. Other estimates taken from *Burbidge and others* [1958].

rate of aggregation during the initial stages of any condensation process, the important short-lived isotopes are U^{236}, Sm^{146}, Pu^{244}, and Cm^{247}. All these have half-lives sufficiently long to have contributed heat during the period 10^7 to 10^8 years after the initial formation. The four short-lived isotopes contributed about 20 times the heat produced by potassium during the 10^8 years. If all this heat was retained by the material, then the temperature increase due to the disintegration of short-lived isotopes was about 3000°C.

The largest known source of energy available for initial heat is the potential energy due to the mutual gravitational attraction of the particles that make up a dust cloud. On aggregation of the particles, this energy is either converted into internal energy of the material by increase of temperature and strain, or it is lost by radiation. *Latimer* [1950] estimated that the accretion of the earth from a dust cloud would have resulted in the production of 4×10^4 joules/gram of material. This is vastly greater than the energy available from radioactivity. It would have been more than sufficient to melt the material of the earth and to establish an initial temperature gradient equal to the melting-point gradient.

It is difficult to estimate the total contribution of gravitational energy because of the uncertainty of the dynamics of the accretion process. Much depends on the temperature attained at the surface of the aggregating body and on the transparency of the surrounding atmosphere to radiation. The kinetic energy of the impinging particle will be converted to thermal energy, but this thermal energy may be immediately reradiated into space. The problem is to estimate the surface temperature obtained during the aggregating process, since this heat could be retained by burial. The short time scale for accretion does not allow a substantial amount of heat to flow from the surface towards the interior.

If the atmosphere of the primitive earth were transparent, the temperature at the accreting surface would be determined by a balance between the energy radiated from the surface and the kinetic energy converted to thermal energy:

$$\rho \frac{GM(r)}{r} \frac{dr}{dt} \approx sT^4 \tag{30}$$

where s is the Stefan-Boltzman constant and $M(r)$ is the mass enclosed in a sphere of radius r. If an aggregation time of 10^8 years for the total mass of the earth is assumed, we find that the surface temperature during the later stages of aggregation was on the order of 700°C.

The temperature of the material within the aggregating earth would have increased because of the adiabatic compression of the material. The adiabatic gradient depends linearly on the thermal expansion of the material. Various estimates of the thermal expansion of the material that makes up the earth have been made [*Birch,* 1938; *Benfield,* 1950). A temperature increase of several hundred degrees due to adiabatic compression appears reasonable, though the data are most uncertain [*Verhoogen,* 1956; *MacDonald,* 1956].

In summary, the long-lived radioactive isotopes probably contributed very little to the initial temperature of the earth. The short-lived isotopes might have contributed enough heat to raise the temperature to 2000° or 3000°C. A conversion of gravitational energy into thermal energy provides a very large source of heat. However, the amount of heat initially trapped depends very critically on the actual mechanics of aggregation and on the conditions pertaining to radiative equilibrium. This is a problem that requires detailed study. On the basis of this unsatisfactory situation we shall use three models for the initial state of the earth. In one model the temperature is assumed to be 4100°C at the center and to decrease outwards along a Simon curve as given in Table 20. In a second model the temperature is assumed to be a constant 1300°C within the earth. This later estimate would perhaps be appropriate for an accreting earth in which equilibrium is maintained between the input of gravitational energy and the output of radiation. At deep levels the adiabatic compression dominates, whereas nearer the surface the converted kinetic energy is the important factor. Finally, an earth initially uniform at 0°C is examined.

TABLE 20—*Initial temperature distributions for hot models*

Depth, km	*T*, °C for models 13, 17, 18, 19, 20	*T*, °C for model 16
0	0	0
300	2050	2672
600	2250	3520
900	2430	4007
1200	2600	4280
1500	2720	4454
2000	2930	4660
2500	3110	4702

Urey and Donn [1956] suggested that the reduction of iron in the asteroidal bodies could be accomplished by free radicals and that large amounts of heat would be produced. The total heat that might be generated in this process depends upon the chemical state of the carbon trapped in the accreting dust. A further source of initial heat is energy stored as radiation damage within the aggregating material. In order for this to be important it is necessary for the material to have been exposed to strong radiation at a stage after its condensation into particles. These last two energy sources are even more speculative than those discussed previously. Both would tend to increase the total amount of heat initially captured within the earth.

Time-dependent solutions—Time-dependent solutions to the heat conduction equation (9) which have a variable conductivity given by equation (26) have been obtained by solving numerically the difference equations (14). The numerical computations were programed for an IBM 704 digital computer and the calculations were carried out at the Massachusetts Institute of Technology Computing Center.

In all the models investigated it was assumed that the material had the bulk composition of chondritic meteorites. The concentrations of uranium and potassium are 1.1×10^{-8} g/g and 8.0×10^{-4} g/g, respectively. The thorium-to-uranium ratio is taken as equal to 4 (see model earth, Table 4). The initial radioactivity is assumed to decay exponentially at a rate fixed by the half-lives of the elements. The parameters common to all models are listed in Table 21. For inhomogeneous models the density distribu-

TABLE 21—*Parameters common to all models*

Parameter		Value
Surface temperature		0°C
Index of refraction		1.7
Energy gap		3.0 ev
Heat capacity		1.3 joules/g °C
Lattice conductivity		0.025 joules/cm sec °C
Decay constants	U^{238}	1.54×10^{-10} years^{-1}
	U^{235}	9.71×10^{-10} years^{-1}
	Th^{232}	0.499×10^{-10} years^{-1}
	K^{40}	5.5×10^{-10} years^{-1}
Heat generation	U	3.07 joules/g year
	Th	0.82 joules/g year
	K	1.12 joules/g year

tion is assumed by Bullen's model 1 [1953, p. 218]. Various initial temperature distributions have been investigated. The temperature distributions for the models of an initially hot earth are listed in Table 20.

The effect of initial temperature distribution on the final temperature is illustrated in Table 22. In models 11, 12, and 13, it is assumed that the core differentiated during the aggregation of the earth and that the radioactivity was initially uniformly distributed throughout the mantle. Models 11 and 12 are initially 'cold,' so that the core is assumed to have formed by some unspecified mechanical separation. The initial temperature in model 13 exceeds the melting temperature of iron, and in this case the core could separate as a fluid phase during the earliest stages of the earth's history.

TABLE 22—*Effect of initial temperature on temperature distribution at 4.5 × 10⁹ years*

Depth, km	Model 11 *T*, °C	Model 12 *T*, °C	Model 13 *T*, °C
100	950	1250	1450
200	1490	1860	2110
400	2040	2520	2900
600	2340	2960	3510
800	2530	3290	4020
1200	2710	3710	4760
1600	2750	3870	5180
2000	2750	4000	5360

Description of models

11. $\epsilon_0 = 10$ cm^{-1}, $\sigma_0 = 10$ ohm^{-1} cm^{-1}; radioactivity uniform in mantle. Temperature at time -4.5×10^9 years = 0°C throughout earth. Final heat flow = 27.6 ergs/cm² sec. Core insulated.
12. $\epsilon_0 = 10$ cm^{-1}, $\sigma_0 = 10$ ohm^{-1} cm^{-1}; radioactivity uniform in mantle. Temperature at time -4.5×10^9 years = 1300°C throughout earth. Final heat flow = 42.9 ergs/cm² sec. Core insulated.
13. $\epsilon_0 = 10$ cm^{-1}, $\sigma_0 = 10$ ohm^{-1} cm^{-1}; radioactivity uniform in mantle. Temperature at time -4.5×10^9 years given in Table 21. Final heat flow 56.3 ergs/cm² sec.

If the initial temperature is uniformly O°C., the temperature in the deep mantle after 4.5 × 10⁹ years is 2700°. The conductivity approaches a constant value of 0.14 joules/cm sec °C. An initial uniform temperature of 1300°C leads to a final temperature of 4000° in the deep mantle, with the conductivity approaching a similar constant value. In this case, the surface heat flow, after 4.5 × 10⁹ years, is about 80 per cent of the observed heat flow. As is illustrated in Figure 3, the temperature curve slightly exceeds the assumed melting curve for diopside in the region between 400 and 900 km. The temperature approaches or exceeds the melting temperature in the region from 100 to 1000 km in a large number of models.

In the hot model 13 the initial temperature distribution is that given in Table 22; the resulting final temperature distribution is shown in Figure 3. The final heat flow is now 56 ergs/cm² sec. This value is in rough agreement with the observed heat flow. Model 13 fails to meet the requirement of a solid mantle, since the final temperature lies well above the estimated melting temperature of diopside at all depths greater than 150 km.

Initial homogeneous earth and the formation of the core—In models 14 and 15 we investigate the thermal history of an initially homogeneous earth having a chondritic composition (Table 23). The initial temperature is assumed to be a uniform 1300°C throughout the earth. Because of the dusty, inhomogeneous character of chondrites, the opacity is assumed to be high in both

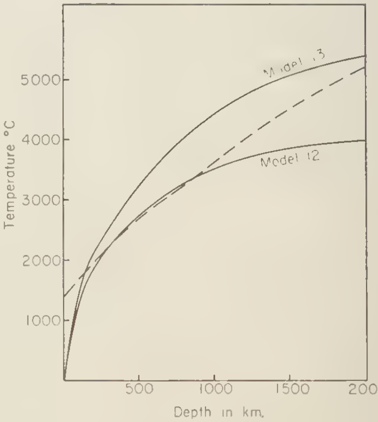


FIG. 3—Temperature distributions in mantle after 4.5 × 10⁹ years. Radioactivity is uniformly distributed in the mantle. In model 12 the initial temperature is a uniform 1300°C (Table 22). In model 13 an initially hot earth is assumed (Table 21 and 22).

TABLE 23—*Temperatures within a homogeneous earth*

Depth, km	Model 14 T, °C	Model 15 T, °C
100	760	720
200	1400	1350
400	2260	2270
600	2710	2780
800	2930	2990
1200	3070	3080
1600	3080	3080
2000	3080	3080
6370	3080	3080

Description of models

14. $\epsilon_0 = 100$ cm^{-1}. $\sigma_0 = 10$ ohm^{-1} cm^{-1}. Radioactivity uniform in earth. Temperature at time -4.5×10^9 years = 1300°C throughout earth. Final heat flow 19.2 ergs/cm^2 sec.
15. $\epsilon_0 = 1000$ cm^{-1}. $\sigma_0 = 10$ ohm^{-1} cm^{-1}. Radioactivity uniform in earth. Temperature at time -4.5×10^9 years = 1300°C throughout earth. Final heat flow 18.0 ergs/cm^2 sec.

models, 100 cm^{-1} for model 14 and 1000 cm^{-1} for model 15. In both cases the final temperature reached deep within the earth is 3081°C. In both models the final heat flow is low; this results from the low initial temperature and the uniform distribution of heat sources.

The development of temperature as a function of time within model 14 is shown in Figures 4 and 5. It is seen that after 0.6×10^9 years the melting point of iron (as estimated by Strong, Fig. 1) is exceeded at a depth of about 400 km. The actual time at which iron would begin to melt depends on the detailed shape of the iron melting curve (Fig. 1). It does seem certain that the temperature would have been raised above the melting point of iron at some time between 0.6 to 2.0×10^9 years after the initial agglomeration of the earth. The temperatures in the mantle nowhere exceed the melting temperature of silicate materials. The thermal development of model 15 is similar to that of model 14, though the time in which the melting temperature of iron is reached is somewhat shorter.

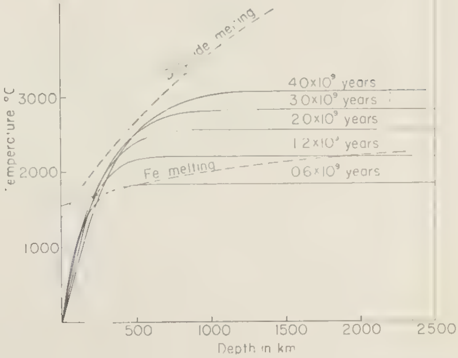


FIG. 4—The development of the temperature distribution within a homogeneous earth of chondritic composition (model 14, see Table 23).

If the earth accreted from chondritic material, the metallic phase would have reached the melting temperature some 10^9 years after the completion of the aggregation process, even if the initial temperature was only 1300°C. This suggests the possibility of a gravitational separation of the core from the mantle some 3 to 4 billion years ago. Such a separation might have completely altered the thermal character of the earth. *Urey* [1952] emphasized that the reorganization of the earth from a uniform, metal-silicate phase mixture to the present state would have released substantial amounts of gravitational energy. The amount of gravitational energy that was converted to heat would depend on the detailed mechanism of the separation. If the process took place at a sufficiently slow rate so that equilibrium was maintained, then the gravitational energy was reversibly converted to internal energy. The increase in temperature was only that due to adiabatic

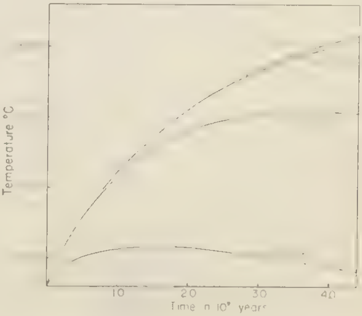


FIG. 5—Development of temperature with time at various depths for model 14.

compression [*Lapwood,* 1952]. An irreversible separation could produce a substantial amount of heat. The order of magnitude of the maximum amount of energy released in going from a homogeneous to a differentiated earth is 10^{31} joules [*Urey,* 1952]. This figure is uncertain, since it is the difference of two large numbers (gravitational energies of a uniform and a differentiated earth). The release of 10^{31} joules provides sufficient energy both to melt a core of iron and to raise the temperature of the whole earth. The surface value for the heat of fusion of iron is 270 joules/g [*Stott and Rendall,* 1953]; so on the order of 5×10^{29} joules are needed to melt the core. The remaining gravitational energy is sufficient to raise the average temperature about 2000°C.

The differentiation of the core would proceed from the upper mantle towards the center of the earth. For such a model the inner, presumably solid, core might represent undifferentiated, primitive material in the form of a solid silicate phase and a fluid metallic phase.

In the following, several models are investigated in which it is assumed that the core differentiated from the mantle at some time after the initial formation of the earth. The differentiation of the core is assumed to have been completed 2.6×10^9 years ago, resulting in the temperature distribution given in Table 21. The time scale for differentiation is thus assumed to be on the order of 10^9 years.

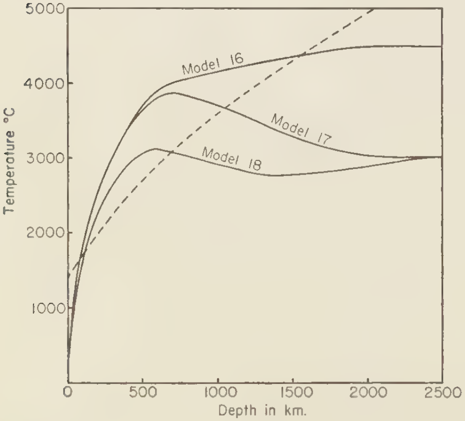


FIG. 6—Temperature distributions in differentiated mantle (Table 24). Models 16 and 17 assume a redistribution of heat sources 2.6×10^9 years ago.

Temperature distributions for differentiated mantle—In models 16, 17, and 18, it is assumed that the radioactivity is concentrated in the upper 600 km of the earth. In all these models the final heat flows are in excess of the observed heat flow. This is the result of the near-surface concentration of the heat sources. A large percentage of the heat produced by radioactivity is reaching the surface, and, in addition, the contribution from initial heat is important.

In model 16 we assume that the radioactivity was concentrated at a time 2.6×10^9 years ago. The temperature distribution at that time is equal to the temperature distribution found in model 13 after 1.8×10^9 years (Table 21). The final temperature distribution is given in Table 24 and in Figure 6. The final temperature exceeds the melting temperature of diopside at depths between 100 and 1500 km.

In model 17 the radioactivity was concentrated during initial formation of the earth. The initial temperature distribution was that given in Table 21. Even though the initial tempera-

TABLE 24—*Temperature distributions for a differentiated mantle*

Depth, km	Model 16 *T*, °C	Model 17 *T*, °C	Model 18 *T*, °C
100	1760	1770	1600
200	2520	2520	2250
400	3420	3390	2900
600	3900	3800	3120
800	4060	3820	3030
1200	4240	3560	2840
1600	4390	3220	2870
2000	4500	3040	2890
2400	4500	3020	3010

Description of models

16. $\epsilon_0 = 10$ cm^{-1}. $\sigma_0 = 10$ ohm^{-1} cm^{-1}. Radioactivity uniform in upper 600 km. Temperature at time -2.6×10^9 years given in Table 21. Final heat flow = 87.4 ergs/cm² sec.
17. $\epsilon_0 = 10$ cm^{-1}. $\sigma_0 = 10$ ohm^{-1} cm^{-1}. Radioactivity uniform in upper 600 km. Temperature at time -4.5×10^9 years given in Table 21. Final heat flow = 87.6 ergs/cm² sec.
18. $\epsilon_0 = 10$ cm^{-1}. $\sigma_0 = 10$ ohm^{-1} cm^{-1}. Radioactivity uniform in upper 600 km. Temperature at time -2.6×10^9 years given in Table 21 Final heat flow = 69.8 ergs/cm² sec.

ture is less than in model 16, the final heat flow is almost identical with that in model 16, as is the near-surface temperature distribution (Fig. 6). The longer time during which the radioactivity has been concentrated near the surface insures that a large percentage of the heat produced in the near-surface region is reaching the surface. The buildup of heat sources near the surface also raises the temperature of the material at the 600-km depth above the temperature at a greater depth within the mantle, and heat is flowing both inwardly and outwardly in this model. The outward flow at a depth of 500 km is 3 times larger than the heat flow toward the center of the earth at a depth of 1000 km.

A lower initial temperature at the time of differentiation is assumed in model 18. In this case the final heat flow and the final temperature distribution are lower than in the case of models 16 and 17. This model also represents a differentiation of the mantle at a time 2.6 × 10^9 years ago, with an initial temperature distribution as given in Table 21. The melting temperature of diopside is exceeded at all depths between 100 and 700 km.

These models illustrate the difficulty with the hypothesis of a differentiated earth in which heat is transported solely by radiative conduction. The process of differentiation will require some initially high temperature within the earth. The heat flow resulting from this initial heat plus the heat flow resulting from the heat sources is larger than that observed. This large heat flow gives a temperature distribution exceeding that of the melting point of silicates. One way to escape from this dilemma is to assume that processes other than conduction are operative in removing the heat. Thus, as the melting point of the silicate material is exceeded, convection and volcanism may become important in removing excess heat. The corresponding thermal history would be complicated. Long intervals in which heat was transported by thermal conduction would be separated by short intervals in which a large amount of heat could be removed from the upper portions of the earth by convection. Only in some such fashion can consistency be obtained between the hypothesis of a chondritic earth, the hypothesis of a highly differentiated mantle, and the observed flow of heat at the surface. An alternate supposition is that the bulk radioactive concentration of the material making up the earth is a factor of 2 less than that of chondritic meteorites.

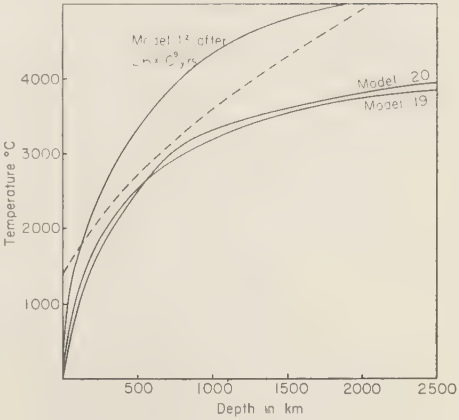


Fig. 7—Temperature distribution for a uniform mantle differentiated from the core 2.6× 10^9 years ago (Table 25). Difference between model 13 after 2.6 × 10^9 years and model 19 is due to variation of rate of radioactive heat production with time.

Miscellaneous models—In model 19 we assume that the core was differentiated at 2.6 × 10^9 years ago and that the radioactivity was uniformly distributed within the mantle with the temperature at this time as given in Table 21. The final heat flow is only 36 ergs/cm² sec as compared with 70 ergs/cm² sec for model 18, in which the radioactivity is concentrated in the upper 600 km. The temperature distribution resulting from a uniform distribution of radioactivity 2.6 × 10^9 years ago is shown in Figure 7 to lie below the melting curve of diopside.

The effect of the variable rate of heat production is illustrated by comparing the temperature distribution in model 13 after 2.6 × 10^9 years with the final temperature distribution in model 19. Both models have the same initial conditions except that model 19 starts 1.9 × 10^9 years later than model 13. The higher initial rate of heat production in model 13 gives rise to a markedly higher temperature, as is shown in Figure 7.

Model 20 is similar to model 19 in all respects except that σ_0 is increased by a factor of 10.

The resulting higher electrical conductivity leads to an initial increase of conductivity, but then the thermal conductivity decreases within the earth. The final temperature distributions are similar (Fig. 7). The heat flow in model 20 is lower than that in model 19. The higher thermal conductivity at depth in model 19 brings more heat to the surface than is the case in model 20. In both cases the final temperature distribution lies below the melting temperature of diopside.

The results for model 19 suggest that it may be possible to construct a model that fits the surface-heat-flow data and the melting-point data but with a substantial proportion of the radioactivity buried at depths greater than a few hundred kilometers.

A distribution of radioactivity which is intermediate between that assumed in model 18 and that in model 19 would give an appropriate heat flow and would satisfy the condition of a solid mantle. In such a case no additional mechanism of heat transport would be needed to disperse excess heat. The critical requirement is the burial of an appreciable amount of the radioactivity at depths below 600 km.

TABLE 25—*Miscellaneous models*

Depth, km	Model 19 T, °C	Model 13 after 2.6×10^9 years T, °C	Model 20 T, °C
100	1130	1540	960
200	1710	2220	1520
400	2330	3020	2220
600	2710	3600	2760
800	2990	4030	3120
1200	3370	4580	3460
1600	3600	4860	3650
2000	3770	5010	3820
2400	3860	...	3940

Description of models

19. $\epsilon_0 = 10 \text{ cm}^{-1}$. $\sigma_0 = 10 \text{ ohm}^{-1} \text{ cm}^{-1}$. Radioactivity uniform in mantle. Temperature at time -2.6×10^9 years given in Table 21.
20. $\epsilon_0 = 10 \text{ cm}^{-1}$. $\sigma_0 = 100 \text{ ohm}^{-1} \text{ cm}^{-1}$. Radioactivity uniform in mantle. Temperature at time -2.6×10^9 years given in Table 21. Final heat flow 28.0 ergs/cm² sec.

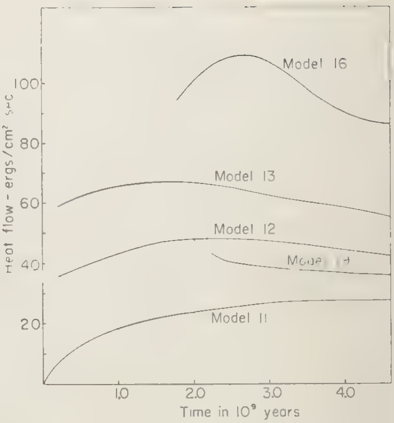


FIG. 8—Variation of surface heat flow with time for various earth models.

History of heat flow—A possible test for any model of the earth's thermal history is the correlation of periods of high volcanic activity with times of high heat flow. Figure 8 shows the history of surface heat flow for a number of models. It is seen that the variations in total heat flow are relatively small. In model 16 there would have been a maximum heat flow some billion years after the differentiation of the core. In this model we would expect a relatively greater portion of volcanic activity to have occurred some 1.5×10^9 years ago. This age is pre-Cambrian, and the geologic evidence is insufficient to confirm or deny the suggestion. In the other models the heat flow remains remarkably uniform, and we should expect no large variations in volcanic activity. This observation suggests a possible way of distinguishing between models in which thermal conduction is the principal mechanism for heat loss and models in which thermal convection plays an important part. If conduction coupled with radiation suffices as the major instrument of heat loss, then volcanic activity and associated phenomena should be uniform throughout the earth's history. On the other hand, short and intense episodes of volcanic activity should be separated by long intervals of quiet if convection resulting from sub-surface melting has been an important element in the transfer of heat. In such a case conduction with radiation would be the principal mechanism of heat transport until the time when the temperature exceeded the melting point gradient. At

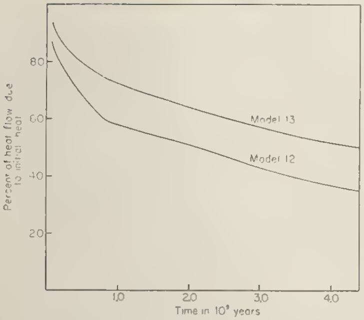


Fig. 9—Relative contribution of initial heat to the total surface heat flow as a function of time for models 12 and 13.

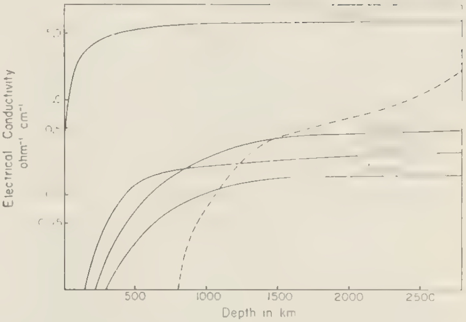


Fig. 10—Distribution of electrical conductivity for various models and for the earth.

this time, then, heat could be transported rapidly by convection and during this time igneous activity should be important. The geologic record is not completely clear. There are suggestions that in the earth's history there have been episodes of great thermal activity separated by periods of quiet. This view, however, is not universally accepted and there are many who would claim that geologic history is characterized by uniformity [*Daly*, 1933; *Gilluly*, 1949].

The importance of the contribution of initial heat to the surface heat flow is illustrated in Figure 9. The proportion of the total surface heat flow that is due to initial heat is plotted as a function of time for models 13 and 12. In model 13 the initial temperature distribution is high and even after 4.5 × 10^9 years 50 per cent of the total surface heat flow is due to the initial heat. In model 12 the initial temperature is a uniform 1300°C, yet this low temperature is sufficient to account for 35 per cent of the total surface heat flow after 4.5 × 10^9 years. In both models all the radioactivity is concentrated in the mantle. The proportion of the total heat flow due to initial heat is somewhat less if the radioactivity is concentrated near the surface.

The contribution of initial heat to the surface heat flow in models 12 and 13 emphasizes the fact that the coincidence of the observed heat flow with the heat production of a chondritic earth is not necessarily indicative of chondritic composition. The earth's material could have a radioactive content half that of chondritic meteorites and make up the total heat flow with appropriate initial conditions. Alternatively, if the earth is chondritic, then the initial heat or a major part of it must have been lost by processes other than ordinary thermal conduction and radiative transfer. This is particularly true if the radioactivity is concentrated towards the surface.

Distribution of electrical conductivity—Figure 10 shows the distribution of electrical conductivity for various models and the estimated distribution of conductivity within the earth. The parameters chosen to represent the effect of electrical conductivity on the thermal conductivity reproduce in a broad way the distribution of electrical conductivity with depth. There is a marked increase in electrical conductivity in the outer few hundred kilometers. This is in agreement with the calculations of Lahiri and Price. In all the models the electrical conductivity tends to have a relatively uniform value within the deep mantle. This again is in rough qualitative agreement with the observed behavior.

It seems likely that no major errors have been included in the calculations by badly estimating the parameter that relates to the electrical properties of the mantle. The temperature distribution is relatively insensitive to variations in the parameters σ_0 and E, as has been noted by *Lubimova* [1958]; (compare models 19 and 20). The electrical conductivity within the earth is not well enough known to allow a choice between the different models. The general features of the distribution of electrical conductivity within the earth could be more closely matched by adjusting the parameters σ_0 and E. The

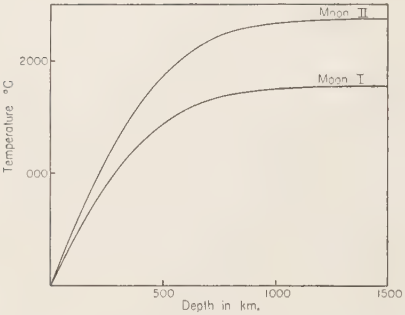


FIG. 11—Temperature distribution in the moon after 4.5 × 10^9 years for a cold moon I and a hot moon II (Table 26).

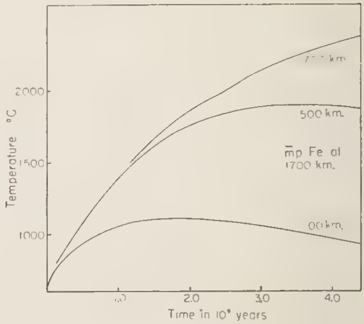


FIG. 13—Development of temperature with time at various depths for moon II.

needed adjustments would not markedly alter the thermal history of a given model.

Thermal history of the moon—Two models of the possible thermal history of the moon have been studied. Model 1 represents a cold moon. The initial temperature is taken as 0°C and the opacity is taken at 100 cm^{-1}, and radiation makes a noticeable contribution to the conductivity. Distribution of temperature after 4.5 billion years is shown in Figure 11 and the development of this temperature as a function of time is shown in Figure 12. It is seen that 3 billion years after the formation of the moon, iron would begin to melt at a depth of 1700 km. At this time there would be a tendency towards the differentiation if a metallic phase were present. Whether such a process would proceed requires detailed investigation, as it is uncertain what the requirements are for formation of a core when melting begins near the center of a small gravitating body. In the cold model of a moon the melting temperature of silicate is not exceeded.

Moon model 2 represents the thermal history of a moon starting with a uniform initial temperature of 600°C and having an opacity of 1000 cm^{-1}. In this case the melting point of iron at a depth of 1700 km would be exceeded at a time 1.6 × 10^9 years after the formation of the moon (Fig. 13). The melting point of iron at 500 km would be exceeded shortly thereafter. Since, in this case, melting of the metallic phase would take place throughout a greater part of

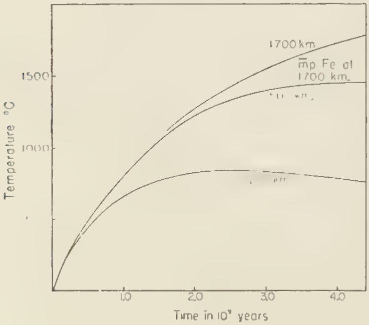


FIG. 12—Development of temperature with time at various depths for moon I.

TABLE 26—*Temperature distribution in moon*

Depth, km	Moon Model 1 T, °C	Moon Model 2 T, °C
100	400	490
200	760	940
300	1050	1320
500	1450	1870
700	1650	2170
900	1730	2310
1100	1770	2360
1300	1780	2370
1500	1780	2380
1700	1780	2380

Description of models

1. $\epsilon_0 = 100 \text{ cm}^{-1}$, $\sigma_0 = 10 \text{ ohm}^{-1} \text{ cm}^{-1}$. Radioactivity uniform throughout moon. Temperature at time -4.5×10^9 years $= 0$°C throughout moon. Final heat flow = 10.3 ergs/cm^2 sec.
2. $\epsilon_0 = 1000 \text{ cm}^{-1}$, $\sigma_0 = 10 \text{ ohm}^{-1} \text{ cm}^{-1}$. Radioactivity uniform throughout moon. Temperature at time -4.5×10^9 years $= 600$°C throughout moon. Final heat flow = 12.4 ergs/cm^2 sec.

the moon, it is difficult to see why a core would not form. The mean moment of inertia of the moon is not known. The moon is not in hydrostatic equilibrium but it is uncertain whether a core could account for this deviation from equilibrium. $(C-A)/A$ for the moon is known; $C - A$ and $C - B$ could be obtained accurately from a lunar orbiting satellite.

Summary

The principal effect of radiative transfer of energy on the distribution of temperature within the earth is the flattening of the temperature gradient with depth. Because of the contribution from radiation, the temperature gradient required to remove heat from the earth's interior decreases as the effective conductivity increases with temperature. As a result, melting temperatures are most closely approached in the outer few hundred kilometers of the earth. If the heat sources are distributed throughout the mantle, the melting point is most closely approached or exceeded at depths of 400 to 800 km. If the radioactivity is concentrated in the upper mantle, the melting point is approached or exceeded in the range of 200 to 600 km. The coincidence of these depths with the seismic low-velocity layer may be significant.

The increase of opacity due to an increased number of free electrons at high temperatures does not influence the temperature distribution in a major way. As a result, the temperature distribution is relatively insensitive to the variation of the parameters that determine the electrical conductivity in the earth. The rapid increase of temperature in the outer few hundred kilometers contributes to the rapid rise of electrical conductivity in this region. The small variation of electrical conductivity at greater depths is consistent with the small variation of temperature in the deep mantle.

The prediction of today's heat flow for a number of models of a chondritic earth having vastly different initial conditions and distributions of heat sources falls, within a factor of 2, of 50 ergs/cm^2 sec. If a factor of 2 is significant, and I assume that it is, then major restrictions can be placed on the thermal history. The contribution of 'initial heat' to present surface heat flow is sufficiently great that, in any model in which the earth has passed through a molten stage, the heat flow is greater than observed heat flow. The concentration of heat sources near the surface further increases the heat flow. A differentiated earth, in which differentiation is the by-product of melting, is inconsistent with the chondritic hypothesis. A differentiated earth, once molten, would require a radioactivity which is at most one-half that of the chondrites. This conclusion is independent of the mode of transport of the heat, unless substantial amounts of heat can be transported by convection at temperatures well below the melting point of the material.

The coincidence of the present surface heat flow with the present rate of heat production in chondritic material can be explained in four ways: 1) The radioactivity which corresponds to a chondritic earth is entirely concentrated within the upper few hundred kilometers. The initial temperature of the earth is less than 1000°C and heat is transported by convection- and conduction-radiation. 2) The chondritic radioactivity is distributed so that about a third of the heat sources lie below 600 km. The average initial temperature is high (1500 to 2000°C). The contribution of initial heat and the higher rate of heat production in the past compensate for heat produced but not reaching the surface. 3) Heat is transported by mass movement of material at temperatures well below the melting point. The excess heat produced in the past can be removed and a differentiated earth is possible. 4) The radioactivity is at most one half that of a chondritic earth. The heat sources are near the surface and the initial temperature can be high.

If the earth were initially a 'cold' homogeneous body of chondritic composition, the metallic phase would have begun to melt at a time 0.5 to 2.0 × 10^9 years after aggregation. The core differentiation could then have taken place at a time long after the initial formation of the earth. The details of the process of the segregation of the core have not been considered, but it is likely that such a process would result in large-scale heating of the earth.

Acknowledgments—The computations described in the paper were carried out at the IBM Computation Center at the Massachusetts Institute of Technology during the spring of 1958. The facilities provided by the Center, in particular the time on the IBM 704, made the study possible.

The manuscript has been critically examined by my colleagues, David Griggs, George Kennedy, Leon Knopoff, and Walter Munk. A large number of their suggestions have been incorporated. I have discussed many of the thermal problems with Sydney Clark. His interest and his knowledge of the field have been most helpful.

References

Adams, J. A. S., Uranium and thorium contents of volcanic rocks, in *Nuclear Geology,* H. Faul, ed., John Wiley & Sons, New York, 89–98, 1954.

Aldrich, L. T., and G. Wetherill, Geochronology by radioactive decay, *Ann. Rev. Nuclear Sci., 8,* 257–298, 1958.

Allan, D. W., Heat in the earth, *Advance. of Sci., 12,* 89–96, 1955.

Altshuler, L. V., K. K. Kupnikov, B. N. Ledenev, V. I. Zhuchikhin, and M. I. Brazhnik, Dynamic compressibility and equation of state of iron under high pressure, *Zhur. Exsptl. i Teoret. Fiz., 34,* 874–885, 1958.

Balchan, A. S., and H. G. Drickamer, The effect of pressure on the spectra of olivine and garnet, *J. Appl. Phys.* (in press), 1959.

Bate, G. L., J. R. Huizenga, and H. A. Potratz, Thorium content of stone meteorites, *Science, 126,* 612–614, 1957.

Bate, G. L., J. R. Huizenga, and H. A. Potratz, Thorium in iron meteorites: a preliminary investigation, *Geochim. et Cosmochim. Acta, 14,* 118–125, 1958.

Benfield, A. E., Temperatures of an accreting earth, *Trans. Am. Geophys. Union, 31,* 53-57, 1950.

Birch, F., Travel times for shear waves in a granite layer, *Bull. Seis. Soc. Am., 28,* 49–56, 1938.

Birch, F., Flow of heat in the front range, Colorado, *Bull. Geol. Soc. Am., 61,* 567–630, 1950.

Birch, F., Recent work on the radioactivity of potassium and some related geophysical problems, *J. Geophys. Research, 56,* 107–126, 1951.

Birch, F., Heat from radioactivity, in *Nuclear Geology,* H. Faul, ed., John Wiley & Sons, New York, 148–174, 1954a.

Birch, F., The present state of geothermal investigations, *Geophysics, 19,* 645–659, 1954b.

Birch, F., Physics of the crust, in *Crust of the Earth,* A. Poldervaart, ed., *Geol. Soc. Am. Spec. Paper 62,* 101–117, 1955.

Birch, F., Differentiation of the mantle, *Bull. Geol. Soc. Am., 69,* 483–486, 1958.

Birch, F., and H. Clark, The thermal conductivity of rocks and its dependence upon temperature and composition, *Am. J. Sci., 238,* 529–558, 613–635, 1940.

Boyd, F. R., and J. England, Melting of diopside under high pressure, *Carnegie Inst. Wash. Year Book, 57,* 173, 1958.

Bullard, E. C., The interior of the earth, in *The Earth as a Planet,* G. Kuiper, ed., University of Chicago Press, 57–137, 1954.

Bullen, K. E., *An Introduction to the Theory of Seismology,* 2nd Ed., Cambridge University Press, 1953.

Burbidge, E. M., G. R. Burbidge, W. Fowler, and F. Hoyle, Synthesis of the elements in stars, *Revs. Modern Phys., 29,* 547–650, 1958.

Carslaw, H. S., and J. C. Jaeger, *Conduction of Heat in Solids,* 2nd Ed., Oxford University Press, 1959.

Clark, S. P., Effect of radiative transfer on temperatures in the earth, *Bull. Geol. Soc. Am., 67,* 1123–1124, 1956.

Clark, S. P., Radiative transfer in the earth's mantle, *Trans. Am. Geophys. Union, 38,* 931–938, 1957a.

Clark, S. P., Absorption spectra of some silicates in the visible and near infrared, *Am. Mineralogist, 42,* 732–742, 1957b.

Clark, S. P., and E. R. Nibelett, Terrestrial heat flow in the Swiss Alps, *Monthly Notices Roy. Astron. Soc., Geophys. Suppl., 7,* 176–195, 1956.

Coster, H. P., The electrical conductivity of rocks at high temperatures, *Monthly Notices Roy. Astron. Soc., Geophys. Suppl., 5,* 193–199, 1948.

Courant, R., K. Friedrichs, and H. Lewy, Uber die portiellen differenzengleichungen der matematischen Physik, *Math. Ann., 100,* 32–74, 1928.

Daly, R. A., *Igneous Rocks and the Depths of the Earth,* McGraw-Hill Book Company, New York, 1933.

Davis, G. H., Radium content of ultramagnetic igneous rocks, III, Meteorites, *Am. J. Sci., 248,* 107–111, 1950.

Davis, G. L., and H. H. Hess, Radium content of ultramafic igneous rocks, II: Geological and chemical implications, *Am. J. Sci., 247,* 856–882, 1949.

Dufay, J., *Galactic Nebulae and Interstellar Matter,* Philosophical Library, New York, 1957.

Edwards, G., and H. C. Urey, Determination of alkali metals in meteorites by a distillation process, *Geochim. et Cosmochim. Acta, 7,* 154–168, 1955.

Eitel, W., Thermal transmissivity by radiation in glass furnaces, *Glass Ind., 36,* 575–581, 592–593, 1955.

Evans, R. D., and C. Goodman, Radioactivity of rocks, *Bull. Geol. Soc. Am., 52.* 459–490, 1941.

Fan, H. Y., Valence semiconductors, germanium and silicon, in *Solid State Physics,* F. Seitz and D. Turnbull, editors, Academic Press, New York, 1955.

Geiss, J., and D. C. Hess, Argon-potassium ages and the isotopic composition of argon from meteorites, *Astrophys. J., 127,* 224–236, 1958.

Gerling, E., and A. Pukanov, Problems of absolute age determination in the Baltic shield, *Geochimiga,* no. 8, 695, 1958.

Gilluly, J., Distribution of mountain building in geologic time, *Bull. Geol. Soc. Am., 60,* 561–590, 1949.

Gilvarry, J. J., The Lindemann and Grüneisen Laws, *Phys. Rev., 102,* 308–316, 1956a.

Gilvarry, J. J., Gruneisen's Law and the fusion curve at high pressure, *Phys. Rev., 102,* 317–325, 1956b.

Gilvarry, J. J., Equation of the fusion curve, *Phys. Rev., 102,* 325–333, 1956c.

Gilvarry, J. J., Grüneisen parameter for a solid under finite stress, *Phys. Rev., 102,* 333–340, 1956d.

Hamaguchi, H., G. W. Reed, and A. Turkevich, Uranium and barium in stone meteorites, *Geochim. et Cosmochim. Acta, 12,* 337–347, 1957.

Hoyle, F., On the condensation of the planets, *Monthly Notices Roy. Astron. Soc., 106,* 406–422, 1946.

Hughes, H., The electrical conductivity of the earth's interior, Ph.D. thesis, University of Cambridge, 1953.

Hughes, H., The conductivity mechanism in the earth's mantle (abstr.), *J. Geophys. Research, 64,* 1108, 1959.

Jacobs, J. A., and D. W. Allan, Temperature and heat flow within the earth, *Trans. Roy. Soc. Can. III, 48,* 33–39, 1954.

Kohman, T., Extinct natural radioactivity: possibilities and potentialities, *Ann. N. Y. Acad. Sci., 62,* 503–542, 1956.

Lahiri, B. N., and A. T. Price, Electromagnetic induction in non-uniform conductors, and the determination of the conductivity of the earth from terrestrial magnetic variations, *Phil. Trans. Roy. Soc. London, Ser. A, 237,* 509–540, 1939.

Lapwood, E. R., The effect of contraction in the cooling by conduction of a gravitating sphere, with special reference to the earth, *Monthly Notices Roy. Astron. Soc., Geophys. Suppl., 6,* 402–407, 1952.

Larsen, E. S., and H. Berman, The microscopic determination of the nonopaque minerals, *U. S. Geol. Survey, Bull. 848,* 1934.

Latimer, W. M., Astrochemical problems in the formation of the earth, *Science, 112,* 101–104, 1950.

Lawson, A. W., and J. C. Jamieson, Energy transfer in the earth's mantle, *J. Geol., 66,* 540–551, 1958.

Lubimova, H. A., Thermal history of the earth with consideration of the variable thermal conductivity of the mantle, *Geophys. J. Roy. Astron. Soc., 1,* 115–134, 1958.

MacDonald, G. J. F., The equation of state of solids at high temperatures and pressures, *J. Geophys. Research, 61,* 387–391, 1956.

MacDonald, G. J. F., Chondrites and the chemical composition of the earth, in *Researches in Geochemistry,* P. H. Abelson, ed., John Wiley & Sons, New York, 476–494, 1959.

Marshall, R. R., Isotopic composition of common leads and continuous differentiation of the crust of the earth from the mantle, *Geochim. et Cosmochim. Acta, 12,* 225–237, 1957.

McDonald, K. L., Penetration of the geomagnetic field through a mantle with variable conductivity, *J. Geophys. Research, 62,* 117–141, 1957.

McQuarrie, M., Thermal conductivity: VI, High temperature method and results for alumina, magnesia, and beryllia from 1000° to 1800°C, *J. Am. Ceram. Soc., 37,* 84–88, 1954a.

McQuarrie, M., Thermal conductivity: VII, Analysis of the variation of conductivity with temperature for Al_2O_3, BeO, and MgO, *J. Am. Ceram. Soc., 37,* 91–95, 1954b.

Patterson, C., Age of meteorites and the earth, *Geochim. et Cosmochim. Acta, 10,* 230–237, 1956.

Peierls, R. E., *Quantum Theory of Solids,* Oxford University Press, 1955.

Reasbeck, P., and K. Mayne, Ages and origin of meteorites, *Nature, 176,* 186–188, 1955.

Reed, G. W., Activation analysis applied to geochemical problems, in *Researches in Geochemistry,* P. H. Abelson, ed., John Wiley & Sons, New York, 458–475, 1959.

Richtmyer, R. D., *Difference Methods for Initial-Value Problems,* Interscience, New York, 1957.

Rikitake, T., Electromagnetic shielding within the earth and geomagnetic secular variation, *Bull. Earthquake Research Inst.,* Tokyo Univ., *29,* 263–270, 1951.

Runcorn, S. K., and D. C. Tozer, The electrical conductivity of olivine at high temperatures and pressures, *Ann. geophys., 11,* 98–102, 1955.

Schumacher, E., Aller Bestimmung von Steinmeteoriten mit der rubidiumstrontium Methode, *Z. Naturforsch., 11a,* 206, 1956.

Senftle, F. E., and N. B. Keevil, Thorium-uranium rations in the theory of genesis of lead ores, *Trans. Am. Geophys. Union, 28,* 732–738, 1947.

Slichter, L. B., Cooling of the earth, *Bull. Geol. Soc. Am., 52,* 561–600, 1941.

Stott, V. H., and J. H. Rendall, The density of molten iron, *J. Iron Steel Inst. London, 175,* 374–378, 1953.

Strong, H. M., The experimental fusion curve of iron to 96,000 atmospheres, *J. Geophys. Research, 64,* 653–659, 1959.

Suchan, H. L., S. Wiederhorn, and H. G. Drickamer, The effect of pressure on the absorption edges of certain elements, *J. Phys. Chem.* (in press), 1959.

Tilton, G., Geochemistry of lead and its parents, *Carnegie Inst. Wash. Year Book, 55,* 167–168, 1956.

Urey, H. C., *The Planets,* Yale University Press, 1952.

Urey, H. C., The cosmic abundances of potassium, uranium, and thorium and the heat balances of the earth, the moon, and Mars, *Proc. Natl. Acad. Sci. U. S., 42,* 889–891, 1956.

Urey, H. C., and H. Craig, The composition of the stone meteorites and the origin of meteorites, *Geochim. et Cosmochim. Acta, 4,* 36–82, 1953.

Urey, H. C., and B. Donn, Chemical heating for meteorites, *Astrophys. J., 124,* 307–310, 1956.

Verhoogen, J., Temperatures within the earth, in

Physics and Chemistry of the Earth, I, Ahrens et al., editors, Pergamon Press, London, 1956.

Von Herzen, R., Heat flow values from southeastern Pacific, *Nature, 183,* 882–883, 1959.

Wasserburg, G. J., and R. J. Hayden, Age of meteorites by the A^{40} K^{40} method, *Phys. Rev., 97,* 86–87, 1955.

Wasserburg, G. J., and R. J. Hayden, Time interval between nucleogenesis and the formation of meteorites, *Nature, 176,* 130, 1955b.

Wilson, J. T., R. D. Russell, and R. M. Farquhar, Radioactivity and age of minerals, in *Handbuch der Physik,* S. Flugge, ed., *47,* Springer, Berlin, 288-363, 1956.

Yoder, H. S., Change of melting point of diopside with pressure, *J. Geol., 60,* 364–374, 1952.

Pressure Solution and the Force of Crystallization—A Phenomenological Theory[1]

PETER K. WEYL

Shell Development Company, Exploration and Production Research Division, Houston, Texas

Abstract—The phenomena of pressure solution and the force of crystallization result from removal or deposition of mineral matter in the region of contact between the mineral grains. We assume that this transport takes place by diffusion in a solution film between the grains. The rate of transport depends on the grain size, the effective normal stress between the grains, the diffusion constant in the solution film, the film thickness, and the stress coefficient of solubility. Values of these parameters required to obtain the amount of pressure solution observed in the St. Peter sandstone are reasonable. The theory predicts an increase in pressure solution with decreasing grain size. The effect of clay films between the grains is also to increase the rate of pressure solution owing to the more rapid rate of diffusion in the clay layer relative to a single solution film between clean mineral grains. This explains how pre-existing clay films in the sediment may, because of the greater rate of pressure solution, develop into stylolites. If the interstitial water is supersaturated with respect to the minerals present, precipitation in the area of contact will take place as long as the supersaturation divided by the stress coefficient of solubility is greater than the average effective normal stress between the grains. If the stress is increased above this limit, pressure solution will take place.

INTRODUCTION

When we look at rocks, we are impressed with their permanence. While other materials with which we are familiar rot and rust, the rocks seem to endure. This static picture, however, is only an illusion created by the disparity between human and geologic time scales. When in our imagination we transcend this time limitation, we see that the drama of the sedimentary rock equals in excitement the more rapid spectacle of life. As in life, the present form of the rock is a result of its development, and to understand this form we must study its history.

The sediments are created by the accumulation of particles by physical and biologic processes. Next follows a slow process of chemical adjustment, in which the less soluble minerals grow at the expense of the more soluble. Meanwhile, the waters squeezed out of the underlying sediment are surging through the forming rock, leaving behind, because of slight changes of solubility, an indelible imprint of their passage. As new sediments are piled on top, stresses that are set up at the grain contacts cause solution, and a gradual shrinkage of the section results. Compaction continues until the porosity has been reduced to zero, at which time a state of minimum gravitational potential energy will be reached.

In order to understand the rocks, we must study the mechanisms that alter them. In the present study we restrict ourselves to the mechanism of solution alteration. Solution alteration is the result of an interaction between the minerals in the rock and the interstitial solutions. The field we have singled out for study is still very broad. We must therefore first subdivide it into a number of different areas, which we shall then attack in turn.

In subdividing our field, we must remember that we are not interested in what is now, but in how it became that way. We must therefore not base our classification on the present appearance, but rather investigate the mechanisms that may have brought it about. Instead of looking at a rock and asking for an explanation of its past, we shall assume a starting material, have a specified process act on it, and observe its development. Three mechanisms of solution alteration suggest themselves. In order to distinguish our mechanistic classification from others based on observation, we shall for simplicity call them mechanisms A, B, and C.

Mechanism A—We define local recrystalliza-

[1] Shell Publication No. 193.

tion of sedimentary material due to differences in mineralogy, crystal size, and degree of crystal purity as mechanism A. This mechanism is operative because of the dynamic equilibrium between the minerals and their ions in the interstitial solution. If slight differences in solubility exist, say, between aragonite and calcite or between calcites of different trace element composition, diffusion in the solution due to the concentration gradient will cause a gradual change by dissolution of the more soluble phase and precipitation of the stabler, less soluble one. Differences in solubility may also be the result of differences in crystal size and crystal purity. If we place a collection of sedimentary materials with the appropriate interstitial solution in a container and maintain it at constant temperature and pressure, neglecting the stress effects of gravity, eventually an equilibrium will be established owing to the action of mechanism A. Only slow processes will be of interest, for the rapid changes will go to completion as the sediment is accumulating.

Mechanism B—Mechanism B is defined as the solution of mineral grains at points of contact and redeposition of this material on free surfaces. This is probably the most important mechanism for the volume reduction of sediments and is dramatically illustrated by the phenomenon of stylolites. In mechanisms A and B, the material locally dissolved is redeposited locally so that no gross change of chemical composition can take place. To move material over large distances, we require mechanism C.

Mechanism C—The cementation or leaching of sediments by moving ground water is classified as mechanism C. In general, the water will move through the rock in solution equilibrium with the minerals making up the bulk of the rock [*Weyl*, 1958], so that solution or precipitation will take place only as the water adjusts itself to solubility changes of the minerals due to changes in temperature and pressure. In the case of accessory minerals, considerable supersaturation may result, which leads to precipitation on nucleation centers or scattered preexisting grains of the mineral.

In the present paper we shall limit our attention to mechanism B. In nature all three mechanisms are operative simultaneously, so that if we wish to apply the results of the present study, we must be careful that the alteration produced has not been significantly affected by the other mechanisms. A more general treatment, taking into account concurrent operation of the three mechanisms, will be possible only after the individual mechanisms have been explored in detail. The present study is thus only a small contribution to the more general study of solution alteration.

The essential aspect of mechanism B is the removal or deposition of mineral matter by solutions in the area of contact between the grains. In formulating the problem, we are thus presupposing that the solutions have access to the area of contact between the grains; in other words, we are assuming that solution films exist between the mineral grains. In the absence of a solution film, we cannot remove material from the contact zone by a solution process, and the theoretical structure to be developed would be without a foundation. We shall therefore start by showing that the existence of such films is indirectly indicated by geologic evidence and by laboratory experiments. Next we shall derive a phenomenological theory based on the existence of such films and show that the implications of the theory are in qualitative agreement with geologic facts. The theory will further allow us to correlate such diverse phenomena as stylolites and the force of crystallization, and indicate how they may be studied in a quantitative way.

A kinetic theory can be developed from first principles involving only the interatomic forces, or we can derive a phenomenologic theory based on certain empirical parameters. We shall limit ourselves to the latter procedure. In order to apply the theory, it will be necessary to estimate the values of these parameters for the particular geologic situation.

The Solution Film

Sedimentary rocks are distinguished from unconsolidated sedimentary materials by a difference in the intergranular contacts. In the rock, the original point contacts between the grains have grown into cohesive areas of contact. This growing together may be due to the introduction of external cement (mechanism C), plastic flow or an alteration of the contact zone. It is the latter case with which we shall concern ourselves in the present paper. Before developing theory based on the assumption that solution ms exist between the grains, we shall first ɔk at some alternate mechanisms that might plain the increase in contact area.

The first mechanism that comes to mind is lume reduction of the sediment due to meanical deformation of the grains. Thus, a sediient made up of rubber balls could be comcted by an elastic deformation of the balls, a diment made up of balls of tar could be comcted by viscous flow of the material, and a diment made up of glass beads could be comcted by crushing. In any one of these cases, e external as well as the internal structure the balls would be deformed. While these echanical mechanisms may be responsible for mpaction in certain cases, they fail to explain orphologies such as those illustrated in the iotomicrographs of Figure 1. These are thin ctions of crinoidal limestones. In Figure 1(*a*) e fossil fragments have been cemented tother (mechanism C), and in Figure 1(*b*) the tergranular contacts are broad and internetrating, indicating that considerable comction has taken place. The internal structure the fossils shows no evidence of deformation it rather indicates that parts of the fossil agments have been dissolved without disturbg the rest. The morphology thus clearly inites that the change cannot have been brought ɔout by a mechanical deformation; it appears be the result of a solution mechanism.

If a solution mechanism is responsible, we ust next inquire whether this mechanism reiires the existence of a continuous solution m or whether dissolution of the contact region n take place if we have true mineral-toineral contacts. A possible mechanism is indited in Figure 2. In step 1 we have two spheril grains with a mineral-to-mineral contact ne. If there is no solution film between the ains, solution can take place only at the argin of the contacts. In step 2, solution at the argins has undercut the contact area. As this ocess continues, the stress on the remaining ineral contact will increase until the strength the mineral is exceeded and crushing occurs tep 3). The crushed material will expose new eas to the solution, and dissolution will conue (step 4) until the grain-to-grain contact restored (step 5) by squeezing out the water. The process is then repeated (step 6), and thus the contact area continues to enlarge.

Without a more detailed analysis, it is difficult to rule out the above process. One does not see evidence of it in thin sections; however, one might argue that the undercut is too small to be visible. The mechanism also fails to explain the phenomenon of the force of crystallization.

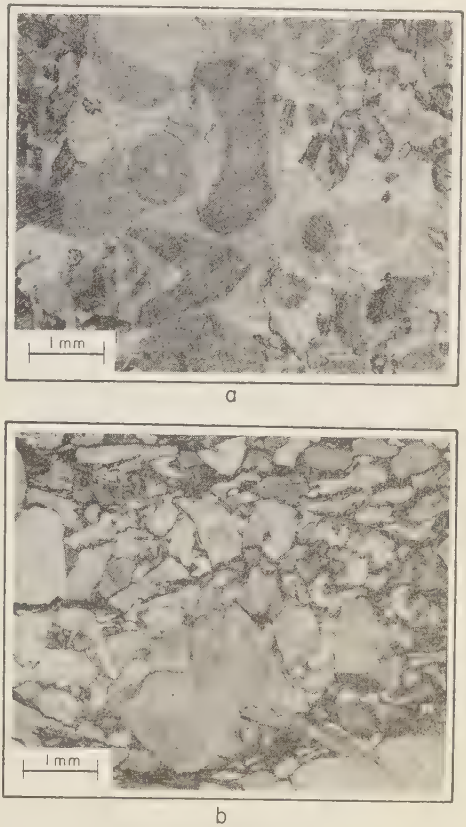


FIG. 1—Thin sections of crinoidal limestone. Bioherm flank beds, Lake Valley formation, Deadman Canyon, Sacramento Mountains, New Mexico. × 20. (Photomicrographs courtesy R. J. Dunham.)

(*a*) Winnowed well-cemented crinoidal limestone exhibiting continuity of cleavage between crinoid columnals and their cementing overgrowths. Small dark particles are fragments of fenestellate bryozoa.

(*b*) Condensed argillaceous crinoidal limestone corroded and compacted along solution seams marked by illite. Note doubly interpenetrating fragment in lower right.

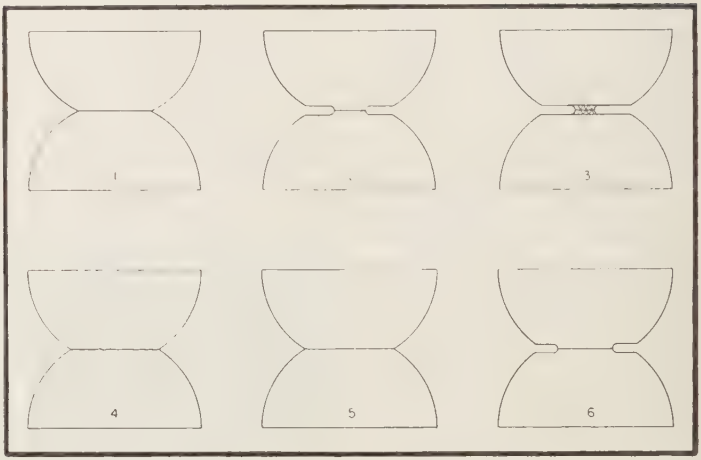


FIG. 2—Pressure solution by stepwise marginal dissolution and crushing.

The fact that crystals can grow by displacing a solid constraint has been demonstrated experimentally by *Becker and Day* [1916] and *Taber* [1916]. In so growing, the crystal exerts a stress, the force of crystallization, against its constraint. This phenomenon, which we shall call the phenomenon of the force of crystallization, is the antithesis of pressure solution where the external stress causes solution of the crystal. The solution can produce precipitation in the region of stress only if a film of solution separates the crystal from its constraint. As long as such a film is required for precipitation, there is no reason why it should not also exist during pressure solution. Owing to the excess of the normal stresses over the hydrostatic pressure, the film must be able to support a shear stress and thus cannot act mechanically as a liquid. In order to be active in solution or precipitation, the ions of the solute must be able to diffuse in this film. While the rate of diffusion does not have to be as rapid as in the free solution, it must be considerably faster than solid diffusion in the mineral grains. In a review paper, *Henniker* [1949] presents a considerable amount of evidence of the existence of solution films having the required properties of mechanical strength and relatively high diffusion rate. We may thus conclude that, although we have not demonstrated the existence of the films, the assumption of their existence is not unreasonable.

Derivation of the General Equations

List of Symbols

A_1, A_2	Integration constants for model A
a	Radius of spherical grain
B_1, B_2	Integration constants for model B
b	Stress coefficient of solubility
c	Concentration of solution (volume mi eral dissolved/volume solution)
c_s	Concentration of saturated solution
Δc	Supersaturation $c - c_s$
D	Diffusion constant of solute in soluti film
d	Density of solid
d_s	Density of solution
d_w	Density of solvent
e	Base of the natural logarithm $e = 2.71$
h	Thickness of solution film
$I(\zeta)$	Compaction integral
J	Diffusion current
K	Apparent compressibility of solute
K_w	Apparent compressibility of solvent
M	Molecular weight
m	Film thickness constant
	Model A $h = m$ equation
	Model B $h = m \exp(-\sigma/\sigma^*)$ equation
p	Hydrostatic pressure
R	Radius of cylinder, gas constant
r	Radial coordinate
S	Total normal stress
T	Absolute temperature
t	Time

Partial molar volume of solute
Partial molar volume of solvent
Potential energy per mole
r/R dimensionless radial coordinate
Value of x where $\sigma(x) = 0$
Fractional thickness of precipitated layer
Vertical coordinate
$\Delta c/\sigma^* b$
Fractional compaction
Radius of contact area
Effective overburden pressure
Effective normal stress
Average effective normal stress
* Stress constant of film (model B)

The objective of this study is the derivation f the dynamic equations which govern the soluion or precipitation of solute in the area of conact between two mineral grains. Intuitively one vould expect that solution or precipitation in he area of contact cannot take place since all he solvent, being a liquid, would be squeezed ut of the contact zone, which would leave a nineral-to-mineral contact. As pointed out preiously, the geologic and laboratory evidence uggests, however, that this conclusion is incorect and that solution and precipitation in the ontact zone do in fact take place. We must herefore conclude that the mineral grains are eparated by a film of solvent which can suport a shear stress but in which the ions of the olute can diffuse at a rate that is rapid comared with the solid diffusion in the mineral rains. Such a film will be characterized by a liffusion constant D and a thickness h which nay depend on the effective normal stress σ cross the film. In the present phenomenological heory, we wish to derive the rate of precipiation or solution in the area of contact in terms f the empirical parameters of the solution and he solvent film. We shall further restrict ourelves to steady-state processes; that is, conlitions such that the thickness of the solvent ilm is independent of time.

In order to obtain a mathematically simple olution, we shall study the solution or precipiation in the area of contact between two coaxial ight circular cylinders of radius R (Figure 3). Ve shall neglect gravity and assume that the ylinders are forced together by a total normal tress S. The hydrostatic pressure of the solution surrounding the cylinders is p, so that the effective normal stress σ is $\sigma = S - p$. The two cylinders are separated by a solvent film which at a distance r from the axis has a thickness $h(r)$. During pressure solution the upper cylinder moves downward with respect to the bottom of the lower cylinder at a rate $-dz/dt$. Inside a radius r, a volume $\pi r^2 dz$ will dissolve in time dt and, in order that the concentration of solute in the film be time-independent, this same amount must diffuse out across the cylindrical surface of area $2\pi rh(r)$. In the case of precipitation (force of crystallization), dz/dt will be positive, and the diffusion must be inward. If c is the volume of dissolved mineral matter per unit volume of solution, the volume

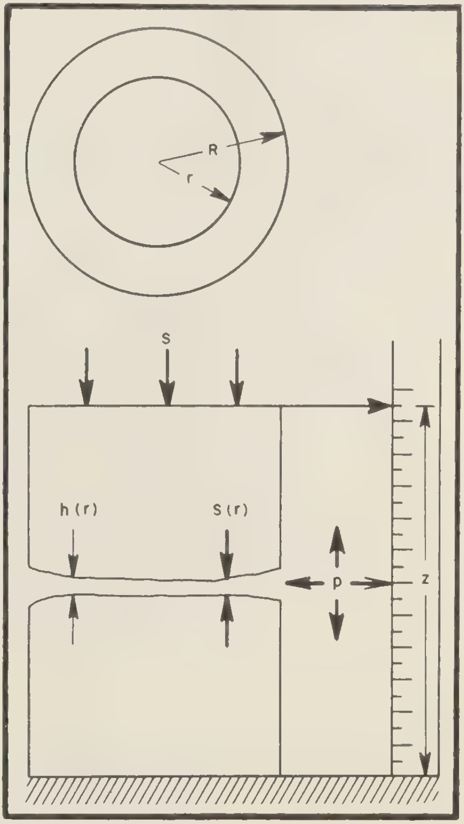


Fig. 3—Pressure solution between two right circular cylinders.

of mineral matter diffusing in unit time is

$$\pi r^2 \frac{dz}{dt} = + 2\pi r h(r)\ D \frac{dc}{dr} \qquad (1)$$

We are here assuming that the diffusion current is proportional to the concentration gradient. This assumption is valid only if the diffusion takes place in the absence of external forces. In the presence of such forces, the theory has to be modified to take into account the forces acting on the diffusing ions. The effects of these forces on the diffusion current are discussed briefly in the Appendix. Assuming these effects to be negligible, we obtain for the rate of compaction

$$-\frac{dz}{dt} = -\frac{2}{r} h(r)\ D \frac{dc}{dr} \qquad (2)$$

For a steady-state process, the rate of compaction $-dz/dt$ must be independent of the radial coordinate. Taking the derivative with respect to r of both sides of equation (2) therefore gives, after division by $-2hD/r$,

$$-\frac{1}{r}\frac{dc}{dr} + \frac{1}{h}\frac{dh}{dr}\frac{dc}{dr} + \frac{1}{D}\frac{dD}{dr}\frac{dc}{dr} + \frac{d^2c}{dr^2} = 0 \quad (3)$$

So far the theory has been quite general. The only assumptions made are that we are dealing with a steady-state process and that the solvent film is thin, so we have essentially a two-dimensional diffusion process. In order to integrate equation (3) it is now necessary to make some assumptions about the dependence of h, c, and D on the effective normal stress $\sigma(r)$ across the solvent film. In the absence of experimental data, we shall assume the following:

1. The diffusion constant D is constant over the entire film. This implies that the term dD/dr is small compared with the other terms in equation (3) and that D can thus be replaced by an average value over the film which, however, can still be a function of the temperature, hydrostatic pressure, and average effective normal stress.

2. The concentration of solute in the film is a linear function of the effective normal stress $\sigma(r)$ across the film; that is, $c = c_o + b\sigma$, where b is the stress coefficient of solubility. c_o and b may depend on the temperature and hydrostatic pressure.

3. The most critical assumption involves
mechanical properties of the solvent film; t
is, the change of thickness of the film with
effective normal stress. In our present ignorar
we shall explore two different models for
film which represent extreme cases. Any conc
sions which are valid for both models will tl
be expected to correspond closely to the r
system. In cases where we obtain widely div
gent results, we may hope that a critical exa
ination of the geologic record may reveal wh
of the models more closely corresponds to re
ity. For the two models we shall assume a fi
thickness independent of the effective norn
stress, and an exponentially decreasing thic
ness.

$$\text{Model A} \qquad h(r) = m$$

$$\frac{dh}{dr} = 0 \qquad ($$

$$\text{Model B} \qquad h = m \exp - \frac{\sigma}{\sigma^*} \qquad ($$

$$\frac{1}{h}\frac{dh}{dr} = -\frac{1}{\sigma^*}\frac{d\sigma}{dr}$$

Model A involves only one empirical constar
the thickness of the film m, whereas model
has as a second constant the stress constant σ
which is the effective normal stress at which tl
thickness of the film is reduced to $1/e$ of its ir
tial value m. Model A is a special case of moc
B with σ^* equal to infinity. In the treatme
that follows, it will however be more convenie
to consider it separately.

Using the above assumptions, we can nc
integrate equation (3) to derive an expressi
for the rate of pressure solution. Using a dime
sionless radial coordinate $x \equiv r/R$ and subs
tuting assumptions 1 and 2 in equation (3),
obtain

$$\frac{d^2\sigma}{dx^2} + \frac{1}{h}\frac{dh}{dx}\frac{d\sigma}{dx} - \frac{1}{x}\frac{d\sigma}{dx} = 0 \qquad ($$

Our assumptions about the mechanical pro
erties of the film (equations 4 and 5) then gi
for models A and B respectively

Model A

$$\frac{d^2\sigma}{dx^2} - \frac{1}{x}\frac{d\sigma}{dx} = 0 \qquad (7a)$$

Model B

$$\frac{d^2\sigma}{dx^2} - \frac{1}{x}\frac{d\sigma}{dx} - \frac{1}{\sigma^*}\left(\frac{d\sigma}{dx}\right)^2 = 0 \qquad (7b)$$

Upon integration we obtain

$$\sigma(x) = A_1 x^2 + A_2 \qquad (8a)$$

$$\sigma(x) = \sigma^* \ln \frac{B_1}{x^2 + B_2} \qquad (8b)$$

where

A_1 and A_2 | B_1 and B_2

re integration constants which depend on the boundary conditions ($x = 1$) and the average ormal stress $\bar{\sigma}$.

The stress gradient $d\sigma/dx = 1/b\ dc/dx$ is

$$\frac{d\sigma}{dx} = 2A_1 x \qquad (9a)$$

$$\frac{d\sigma}{dx} = -2\sigma^* \frac{x}{x^2 + B_2} \qquad (9b)$$

and the thickness of the solvent film is

$$h = m \qquad (10a)$$

$$h = m\frac{x^2 + B_2}{B_1} \qquad (10b)$$

giving for the rate of compaction $-dz/dt$ from equation (2)

$$-\frac{dz}{dt} = -4Dmb\frac{A_1}{R^2} \qquad (11a)$$

$$-\frac{dz}{dt} = 4Dmb\frac{\sigma^*}{R^2 B_1} \qquad (11b)$$

Next we must investigate the range of the ntegration constants that give physically meanngful solutions to the differential equation. The limensionless radial variable x ranges from) to $+1$. For this range in x, the effective noraal stress must remain finite. As the overburden ressure is always larger than the hydrostatic ressure, the effective normal stress must be ositive at least over part of the range of x.

Model A—For this model the effective noraal stress will be finite as long as the integration onstants A_1 and A_2 are finite. Figure 4 is a plot f the A_1–A_2 plane indicating the various regions f interest. They are the following:

Region	$d\sigma/dx$	$\sigma(0)$	$\sigma(1)$
I	negative	positive	positive
II	negative	positive	zero
III	negative	positive	negative
IV	positive	positive	positive
V	positive	zero	positive
VI	positive	negative	positive

Model B—From equation (8b) we see that the effective normal stress becomes infinite for some x between 0 and $+1$ if the integration constant B_2 is between 0 and -1. This region

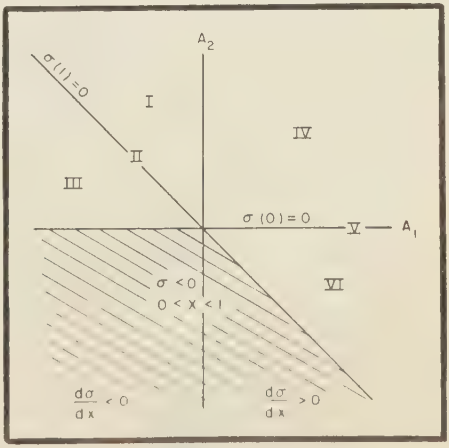


Fig. 4—A_1-A_2 plane indicating the regions of the integration constants for model A.

of the B_1–B_2 plane therefore does not lead to physically meaningful solutions. The regions corresponding to I through VI of model A are indicated in Figure 5. $d\sigma/dx$ is positive if B_2 is less than -1 and is negative if B_2 is positive. Typical curves of the effective normal stress as a function of the radial coordinate x are plotted in Figure 6 for the six regions for both models A and B.

Pressure Solution

Having derived the general equations for transport of mineral matter by diffusion in a solvent film, we are now ready to apply the equations to the phenomenon of pressure solution. In order that pressure solution be possible, it is necessary that the mineral be more soluble when under stress. The stress coefficient of solubility b must therefore be positive. As the diffusion has to be outwards in order to remove material, the stress derivative with respect to the radial coordinate must be negative; that is, the stress must decrease as we approach the circumference. In general, the solution outside the solvent film will be saturated with respect to the external faces of the mineral [*Weyl*, 1958]. As the normal stress on these faces is the hydrostatic pressure p, the effective normal stress is zero. The effective normal stress at the boundary of the film ($x = 1$) must therefore vanish [$\sigma(1) = 0$]. The above conditions are satisfied only for region II of the general equation. In summary, the conditions for pressure solution are

$$b > 0 \qquad (12a)$$

$$\frac{d\sigma}{dx} < 0 \qquad (12b)$$

$$\sigma(1) = 0 \qquad (12c)$$

Pressure solution is also possible in undersaturated and supersaturated solutions (regions I and III, respectively). We shall defer discussion of these cases for the present.

Model A—In order to satisfy the conditions for pressure solution (equations 12), A_2 must be positive and equal to $-A_1$. The effective normal stress is therefore

$$\sigma(x) = -A_1(1 - x^2) \qquad (13)$$

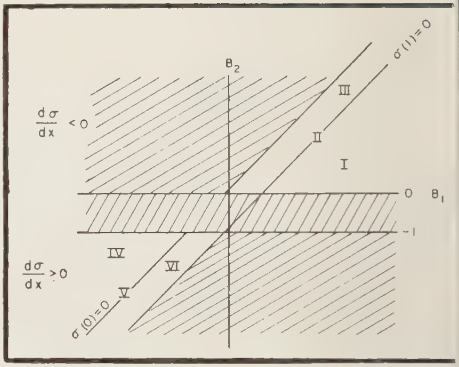


Fig. 5—B_1-B_2 plane indicating the regions of th integration constants for model B.

The average effective normal stress $\bar{\sigma}$ obtaine by integrating over the circular area is

$$\bar{\sigma} = \frac{\int_0^1 -2\pi x A_1(1 - x^2)\,dx}{\pi} = -\tfrac{1}{2} A_1 \quad (14$$

The rate of compaction is therefore

$$-\frac{dz}{dt} = 8Dmb\frac{\bar{\sigma}}{R^2} \qquad (15$$

Model B—In order to satisfy the condition of pressure solution (equations 12),

$$\sigma(1) = 0 = \sigma^* \ln\frac{B_1}{1 + B_2} \qquad (16$$

Therefore

$$B_1 = 1 + B_2 \qquad (17$$

and

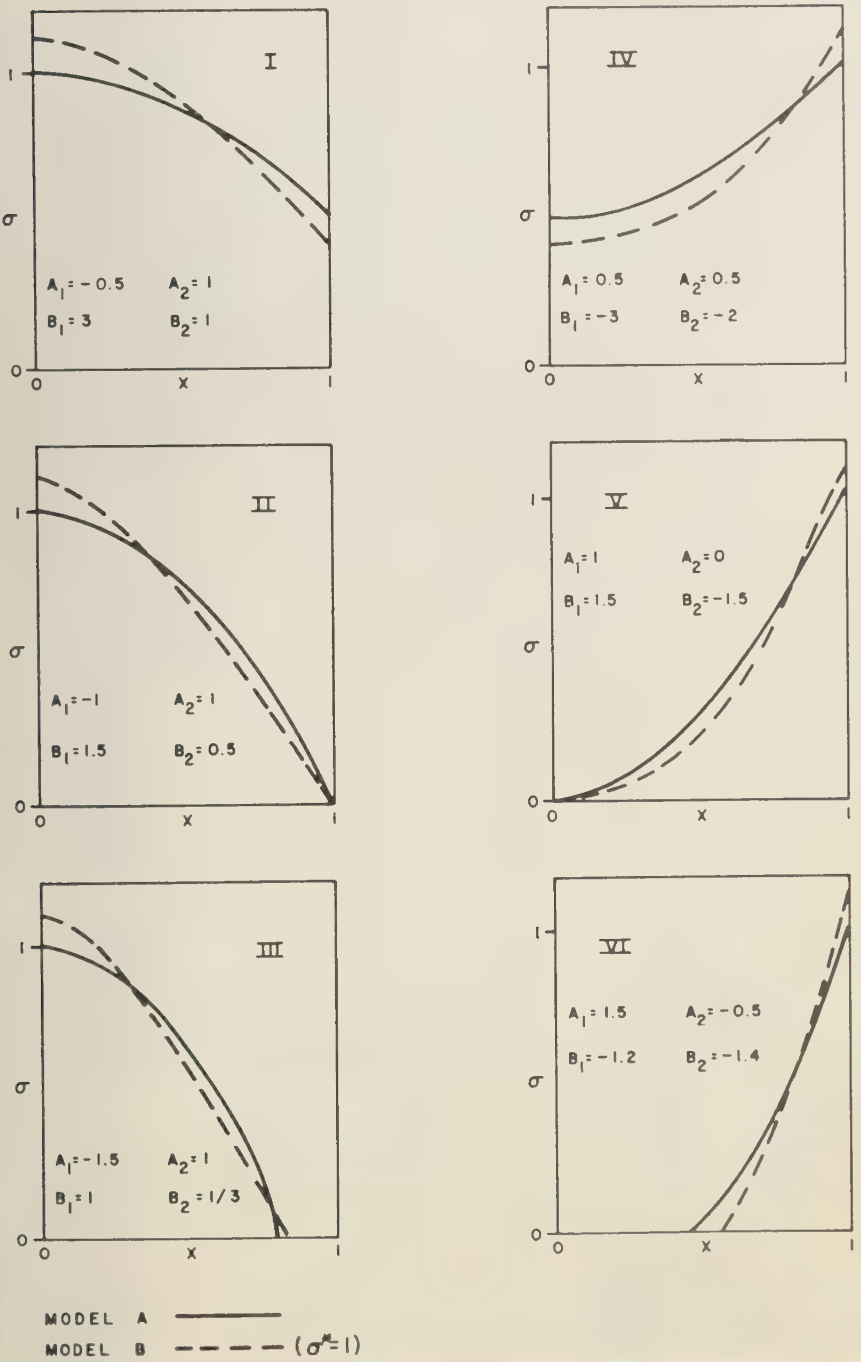
$$\sigma(x) = \sigma^* \ln\frac{B_1}{x^2 + B_1 - 1} \qquad (18$$

The average effective normal stress $\bar{\sigma}$ is

$$\bar{\sigma} = \sigma^* \int_0^1 2x \ln\frac{B_1}{x^2 + B_1 - 1}\,dx \qquad (19$$

$$\frac{\bar{\sigma}}{\sigma^*} = 1 - (B_1 - 1)\ln\frac{B_1}{B_1 - 1}$$

Since for region II, B_1 is larger than 1 (Figur 5), equation (19) can be expanded in power of $1/B_1$.



IG. 6—Typical curves of the effective normal stress as a function of the radial coordinate x for the six regions.

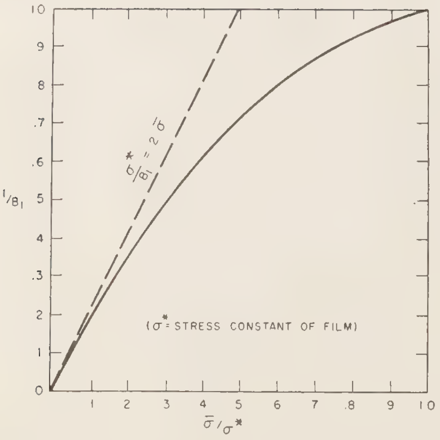


Fig. 7—$1/B_1$ as a function of the average effective normal stress for model B.

$$\frac{\bar{\sigma}}{\sigma^*} = \frac{1}{1 \cdot 2B_1} + \frac{1}{2 \cdot 3B_1^2} \cdots, + \frac{1}{n(n+1)B_1^n} \tag{20}$$

For small values of the average effective stress —that is, values of $\bar{\sigma}$ small compared to σ^*— we therefore find, as we should expect, that the rate of compaction is the same as for model A and is given by equation (15). Figure 7 is a plot of $1/B_1$ as a function of the average effective stress. As it approaches the stress constant of the film, the stress at the center of the film approaches infinity. Finally, if $\bar{\sigma} = \sigma^*$, the filn is broken for this particular model and pressur solution should cease.

General discussion, pressure solution betwee spheres—From equation (15) we see that th rate of compaction is directly proportional t the average effective stress and inversely proportional to the square of the radius of th contact area. If the thickness of the solvent filn does not remain constant but decreases as th stress is increased, the rate of compaction wi increase at a rate slower than linear with th average stress. In order to get a geologically illustrative picture of the compaction due to pressure solution, we shall investigate the followin model. Consider a rock made up of identica spherical grains of unit radius, packed in simple cubic lattice (Fig. 8). Since the maximum principal stress is in the vertical directior we shall assume that pressure solution take place only at the horizontal contacts betwee the spheres. (In general, pressure solution wi not be limited to these contacts but will b more rapid in the direction of the maximur principal stress.) As pressure solution take place, the unit cell of the pack will be reduce in height from an initial value of 2 to $2(1 - \zeta)$ where ζ is the fractional compaction. The material dissolved will be precipitated on the extern

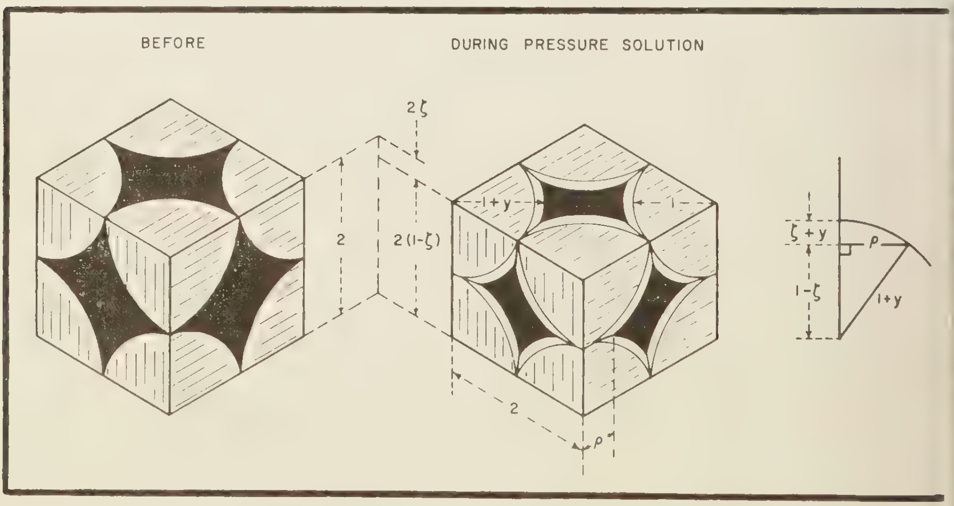


Fig. 8—Pressure solution of identical spheres of simple cubic packing.

urfaces of the grains. For simplicity we shall ssume that the precipitate forms a concentric pherical shell of thickness y. The unit sphere vill thus be transformed into a larger sphere of adius $1 + y$, truncated by the six planes of the init cell. The four lateral truncations will have height y, and the two truncations at the top nd bottom will have a height $\zeta + y$. Since the olume of a segment of a sphere of radius r vith height h is given by

$$\text{Volume} = \tfrac{1}{3}\pi h^2(3r - h)$$

onservation of volume requires the following elationship between ζ and y:

$$\pi = \tfrac{4}{3}\pi(1 + y)^3$$
$$- \tfrac{2}{3}\pi(\zeta + y)^2(3 + 2y - \zeta) - \tfrac{4}{3}\pi y^2(3 + 2y) \qquad (21)$$

he radius of the contact area ρ is given by Fig. 8)

$$\rho^2 = (1 + y)^2 - (1 - \zeta)^2 \qquad (22)$$

Equation (21) is valid only as long as the adius of the contact zone ρ is less than unity. t $\rho = 1$, $y = 0.160$ and $\zeta = 0.412$. At this mount of pressure solution, the initial porosity f 47.7 per cent has been reduced to 10.9 per ent. Generalizing from the unit sphere to a phere of radius a, we note that ζ is still the ractional compaction; that is, the change in hickness of the sediment per unit initial thickness. The radius of the contact zone is now given y $a\rho$. We define the effective overburden ressure Σ acting on the aggregate as a whole s the weight of overburden (including the nterstitial fluids) per unit area, minus the fluid ressure. This pressure will be supported by $/4a^2$ grains. The average effective normal stress n the contact zone is therefore

$$\bar{\sigma} = \frac{4a^2\Sigma}{\pi a^2\rho^2} = \frac{4\Sigma}{\pi\rho^2} \qquad (23)$$

ssuming that model A is a good approximation, we then obtain from equation (15) the ime rate of change of compaction:

$$a\frac{d\zeta}{dt} = \frac{8Dmb\bar{\sigma}}{a^2\rho^2} = \frac{32Dmb\Sigma}{\pi a^2\rho^4}$$

Therefore

$$\frac{d\zeta}{dt} = \frac{32DmbΣ}{\pi a^3\rho^4} \qquad (24)$$

The total fractional compaction is obtained by integrating (24) with respect to time. When integrating, we must remember that ρ is a function of ζ through (21) and (22), and that D and b may be a function of the amount of burial. Separating variables and integrating, we obtain

$$\int_0^{\zeta} \rho^4\, dz \equiv I(\zeta) = \frac{32m}{\pi a^3}\int_0^{t} Db\Sigma d\tau \qquad (25)$$

The compaction integral $I(\zeta)$ has been evaluated numerically and is tabulated in Table 1 and plotted in Figure 9. If the diffusion constant and the pressure coefficient of solubility are assumed to remain constant, the time required for a given fractional compaction ζ under conditions of constant and linearly increasing effective overburden stress is

TABLE 1—*Pressure solution of spheres*

Fractional compaction ζ	Porosity %	Porosity / Initial porosity	$I(\zeta)$	$\sqrt{I(\zeta)}$
0.00	47.7	1.000	.000000	.0000
.05	44.9	.941	.000167	.0130
.10	41.8	.876	.00134	.0366
.15	38.4	.805	.00455	.0677
.20	34.6	.721	.0109	.1044
.25	30.2	.633	.0217	.1474
.30	25.2	.528	.0383	.1965
.35	19.5	.408	.0631	.251
.40	12.7	.266	.0997	.318

constant overburden

$$t = 0.0994\,\frac{a^3 I(\zeta)}{Dmb\Sigma} \qquad (26a)$$

linearly increasing overburden

$$t = 0.443\sqrt{\frac{a^3 I(\zeta)}{Dmb(d\Sigma/dt)}} \qquad (26b)$$

In the general case, the amount of pressure solution will depend on the time-overburden history of the sediment. The age and present depth

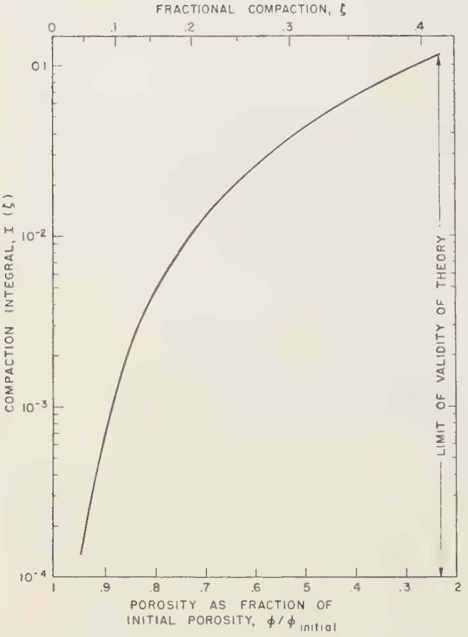


Fig. 9—Compaction integral $I(\zeta)$ for spheres in simple cubic packing.

of burial are not sufficient criteria for estimating the amount of pressure solution. Rather the overburden-time integral (equation 25) must be estimated from the geologic history of the sediment.

Application of the Theory of Pressure Solution to the St. Peter Sandstone

In order to illustrate the development c pressure solution during the history of a sediment, we shall consider the St. Peter sandston in St. Charles County, Missouri, which has bee investigated by *Heald* [1956]. The reconstruction of the overburden-time curve is at best difficult task. It is not our purpose in the present paper either to discuss how these curve may be estimated or to justify the particula curve we shall use. Figure 10 is an approximat overburden-time curve for the St. Peter draw from data kindly supplied by R. C. Verno (private communication). The National Research Council time scale was used, and roug guesses were made as to the erosional histor We present this curve without any claim that is the best possible estimate but merely to illustrate how the history of pressure solutio may be reconstructed if the overburden-tim curve is known.

Using the curve of Figure 10, we must no integrate (25) with time to obtain the tim dependence of the integral $I(\zeta)$. First we sha assume that D and b, the diffusion constant an the stress coefficient of solubility, are essentiall constant independent of depth. In this case, w merely integrate Figure 10 with respect to tim This integral is the solid line in Figure 11. Ov the entire geological history, this integr

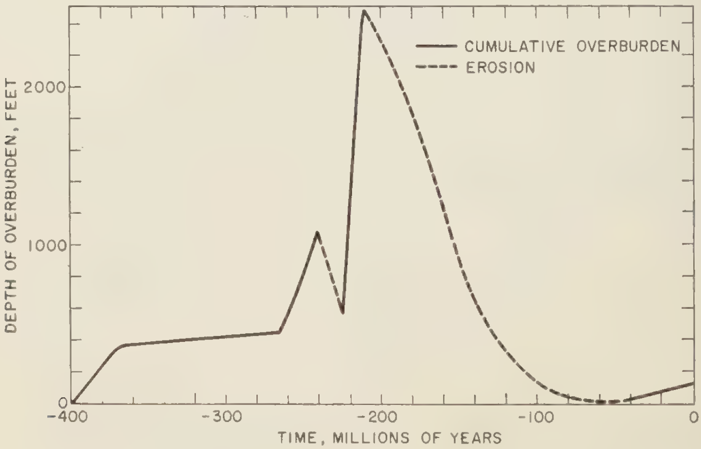


Fig. 10—Approximate overburden-time curve for St. Peter sandstone, St. Charles Co., Missouri.

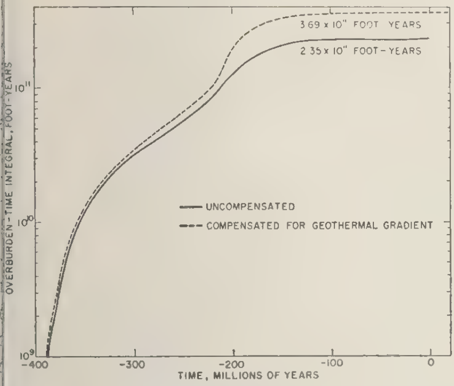


Fig. 11—Overburden-time integral, St. Peter sandstone, St. Charles Co., Missouri.

amounts to 2.35 × 10¹¹ foot-years. According to *Heald* [1956], the present porosity of this sand in the clay-free region (specimen 22) is 14 per cent. If an initial porosity of 37 per cent is assumed, the relative porosity—that is, the present porosity divided by the initial porosity —is 0.38. Assuming that our theoretical model of spheres in cubic packing is a reasonable approximation, we can then reconstruct the porosity history by using the numerical data of the compaction integral (Fig. 9). In fitting these data to the assumed initial and present porosity, we obtain an empirical constant for the product *Dmb*. The porosity history is plotted as the solid line in Figure 12. We find that the porosity decreased at a fairly uniform rate from the time of deposition (Ordovician) until about 170 million years ago (mid-Triassic) and then remained constant.

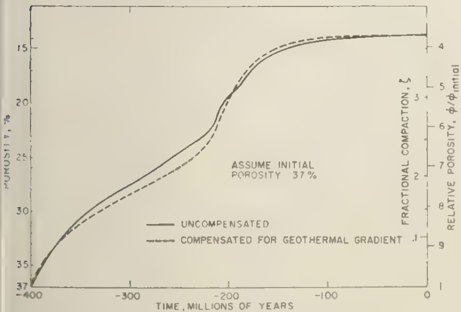


Fig. 12—History of pressure solution of St. Peter sandstone.

In this first approximation we have assumed that the diffusion constant and the stress coefficient of solubility of quartz remain constant independent of the depth of burial. We shall now refine this solution by correcting for the temperature variation of these parameters. If we assume a geothermal gradient of 10°C per thousand feet, at maximum burial the temperature will have increased by almost 25°C. Over this interval the diffusion constant increases by a factor of 1.95, the ratio being approximately a straight line on a log ratio-vs.-temperature plot. In the absence of information on the change of solubility with stress in the solvent film *b*, we shall assume that *b* is equivalent to the pressure coefficient of solubility in the free solution. No data on the temperature variation of the pressure coefficient of solubility of quartz are available for the temperatures of interest. We shall therefore assume that the fractional pressure coefficient of solubility is essentially a constant, so that the pressure coefficient is proportional to the solubility. By extrapolating the high-temperature solubility data of *Ken-*

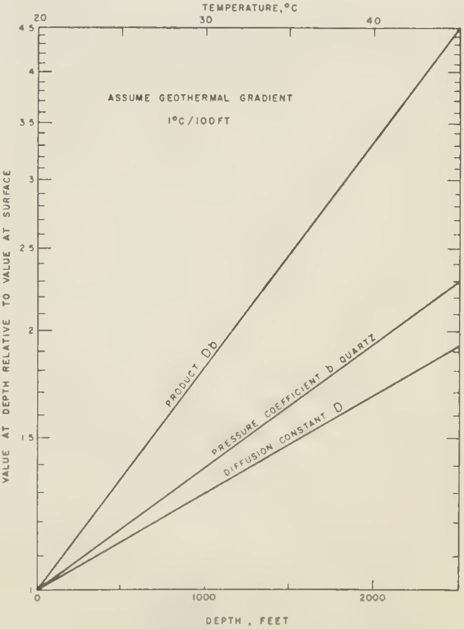


Fig. 13—Temperature correction for quartz (rough estimate).

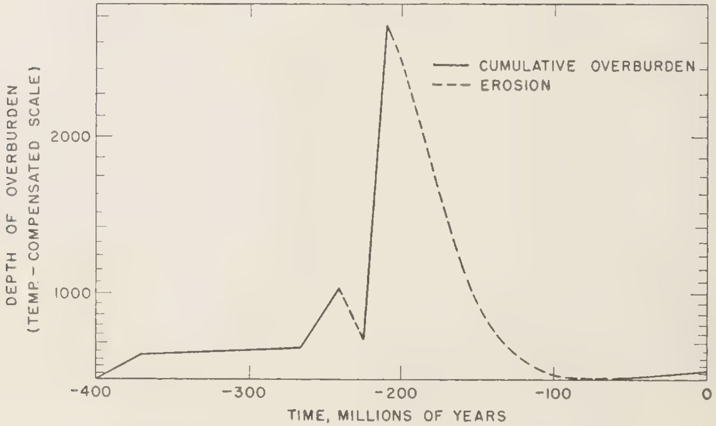


FIG. 14—Approximate temperature-compensated overburden-time curve for St. Peter sandstone, St. Charles Co., Missouri.

nedy [1950], we find that the solubility increases approximately by a factor of 2.3 for a 25°C temperature rise. Drawing this also as a straight line on the semilog plot (Fig. 13), we can obtain an approximate value for the product Db relative to its value at the surface as a function of the depth. This curve is of course highly speculative and is presented only to illustrate how a temperature correction may be made and to indicate the order of magnitude of this correction. We see that for our assumption the product Db has increased by a factor of 4.5 at 2500 feet.

Using the ratio of the product Db at depth relative to its value at the surface, we now construct a temperature-compensated depth scale. On this scale, 1 foot at 2500 feet burial is equivalent to 4.5 feet at zero burial. Figure 14 is the overburden-time curve replotted on a compensated depth scale. By integrating this curve, we obtain the dotted curve of Figure 11. The integral up to the present is 3.69 × 10^{11} compared with 2.35 × 10^{11} for the uncompensated curve. Using the compensated integral, we can then reconstruct the porosity history which is plotted as the dotted curve of Figure 12. We see that in the case of the St. Peter sandstone, the temperature correction changes the porosity history relatively little.

Having worked out the approximate porosity history of the St. Peter sandstone, we can use the data to get an order-of-magnitude estimate of the product Dmb for quartz. The overburden-depth integral was approximately 3 × 10^{11} foot-years. Assuming an effective overburden pressure Σ of 1 atmosphere per 30 feet of burial, we obtain an effective overburden-time integral of 10^{10} atmosphere-years. For a ratio of 0.38 of the present to initial porosity, the compaction integral $I(\zeta)$ is 0.082. The average grain diameter of the St. Peter sandstone (Heald's specimen 22) is about 0.4 mm. Substituting these values in equation (25), we find

$$I(\zeta) = 0.082$$

$$= \frac{32m(\text{cm})\, D\, \dfrac{\text{cm}^2}{\text{yr}}\, b(\text{atm}^{-1}) \times 10^{10}(\text{atm-yr})}{\pi \times 8 \times 10^{-6}(\text{cm}^3)} \qquad (27)$$

Therefore

$$Dmb = 6.4 \times 10^{-18}\left(\frac{\text{cm}^3}{\text{atm-yr}}\right)$$

In order to see if this value is reasonable, we shall make some guesses as to the values of the film thickness m and the pressure coefficient of solubility b and calculate the value of the diffusion constant in the solution film D.

The thickness of the solution film m—The solution film probably amounts to a few atomic layers of water. We shall therefore assume a value of 10^{-7} cm for its thickness.

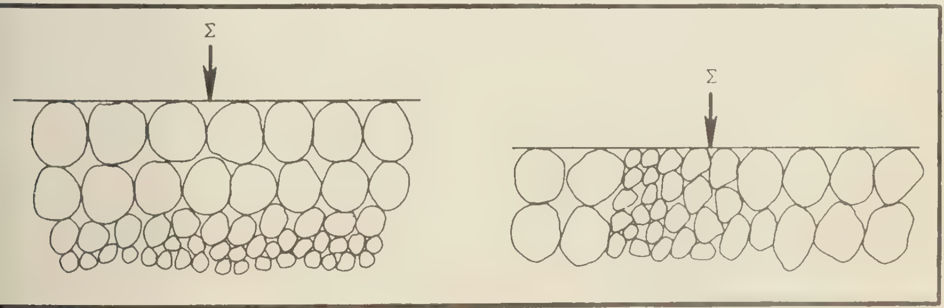
The stress coefficient of solubility b—In the osence of direct data, we shall assume that the nange of solubility in the film with normal ress is of the same order as the change of olubility with hydrostatic pressure in the free olution. We shall further assume that the ressure coefficient of solubility is proportional) the solubility. This is equivalent to assuming nat the fractional change of solubility with ressure is constant. The data of *Kennedy* 1950] at higher temperatures indicate that this ractional coefficient of solubility with pressure of the order of 2 × 10^{-4} atm^{-1} for quartz. ssuming a solubility of 10 ppm, we obtain a alue for b of 6 × 10^{-10} cm^3 quartz/cm^3 solu-on, atmosphere.

Substituting these values of m and b, we find value of 0.11 cm^2/year for the diffusion con-ant in the solution film. In free solution the iffusion constant is about 10^{-5} cm^2/sec, or 316 m^2/year. Thus the rate of diffusion is approxi-ately 3000 times less in the solution film than free water. In the absence of data, all we can ggest is that this value is not unreasonable. ur rough calculations therefore indicate that lthough we have not proved that the model is orrect, we do obtain geologically reasonable re-ults.

The Effect of Grain Size on the Rate of Pressure Solution

From equation (25) we see that if the other arameters are the same, the compaction inte-ral $I(\zeta)$ varies inversely as the cube of the grain radius. In an actual sediment, the grain size will vary from place to place. In studying the effect of a variation of grain size on the rate of pressure solution, we must distinguish between a grain-size variation in the direction, and perpendicular to the direction, of the maximum principal stress. These two idealized cases are illustrated in Figure 15. If the grain-size variation is in the direction of the maximum principal stress (Fig. 15*a*) for adjacent layers, the product of the compaction integral and the cube of the grain size must be a constant. Figure 16 is a plot of fractional compaction vs. relative grain size, on which lines of constant $I(\zeta)a^3$ have been drawn. Thus if the largest grains retain 0.9 of their initial porosity, grains of one-half this diameter will have their porosity reduced to 0.79 of the initial value, and grains of one-quarter the diameter will retain only 0.53 of their initial porosity.

If the grain-size variation is perpendicular to the direction of the maximum principal stress (Fig. 15*b*), the situation is more complicated. If compaction is to take place uniformly, a readjustment of the effective overburden pressure carried by the coarse and fine grains must take place so that the rates of pressure solution are equalized. In this case, the product of the effective overburden pressure times the cube of the grain size must be constant. A random mixture of grain sizes (poor sorting) will lead to preferential pressure solution of the finer grains accompanied by a readjustment of the positions of the coarse grains.



ig. 15—Relationships between the grain-size variation and the direction of the maximum principal stress.

a) Grain-size variation in the direction of the maximum principal stress.

(*b*) Grain-size variation perpendicular to the direction of the maximum principal stress.

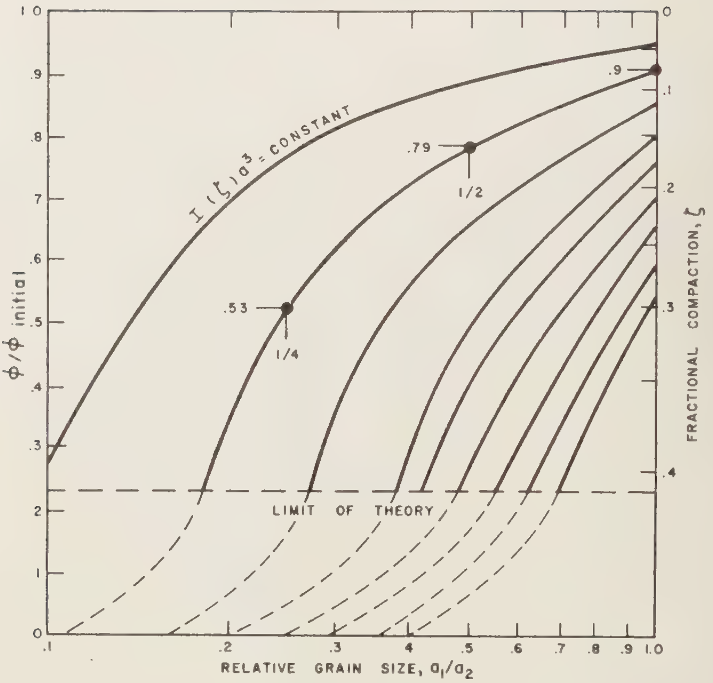


Fig. 16—Effect of grain size on pressure solution. Grain-size variation in the direction of the maximum principal stress.

The effect of grain size on pressure solution is well illustrated by Figure 17, which is Plate 2A of *Heald* [1956]. His caption reads "*A*, shows variation in pore-space reduction due to differences in grain size. Porosity in fine-grained bed near bottom of photograph has been nearly eliminated by pressure solution, whereas considerable pore space remains in coarse bed above."

The Effect of Clay on Pressure Solution—Stylolites

In the foregoing discussion we have assumed that pressure solution takes place between clean mineral grains. In nature, the grains of the clastic sediments may often be coated with clay or contain clay partings. Geologic evidence indicates that this clay has a marked effect on the rate of pressure solution. Thus, *Heald* [1956], after an examination of pressure solution of the Simpson and St. Peter sandstones, wrote "Clay must in some way promote pressure solution because the differences in amount of solution could hardly be due to variations in pressure, temperature, or circulating waters over distances which are no more than a few millimeters in some specimens. . . . The role of clay seems to be similar to that of a catalyst. Pressure solution may occur without clay, but if other conditions are favorable, clay accelerates the process."

Stylolites are one of the most dramatic illustrations of pressure solution. They result from pressure solution on both sides of an undulating surface, the amounts of solution above and below the surface varying while their sum remains the same. The surface usually contains clay or other fine-grained material. Whether this clay film is cause or effect is not clear from the published literature. Thus *Stockdale* [1922] believed the clay to be 'the solution residue of the dissolved lime-mass.' *Heald* [1956], on the other hand, wrote 'Although most of the clay on seams may represent a residue, original clay partings could have served as loci for stylolitic solution.' In view of the above, it becomes important to

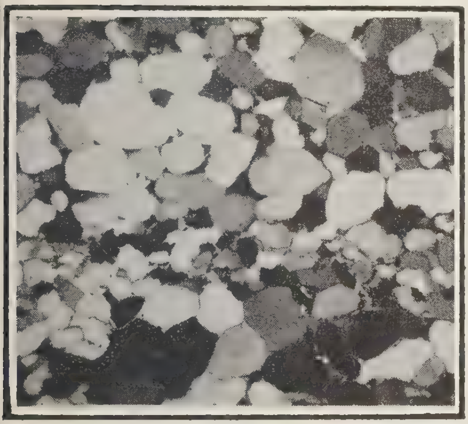


FIG. 17—Effect of grain size on pressure solution. St. Peter sandstone, Newton Co., Arkansas. Crossed nicols, ×30. (After *Heald* [1956]—reprinted by permission of the *Journal of Geology* and the University of Chicago Press.)

inquire how our model has to be modified if the solution film between the mineral grains is replaced by a layer of clay. The significance of the solution film is that it affords a path for diffusion of the pressure-dissolved solute. We must therefore determine how this rate of diffusion is altered by the presence of a clay film.

The clay film consists of a collection of clay platelets with associated water films. The thickness of each platelet with its water film is of the order of 20 A. A 0.01-mm clay film between two mineral grains will thus contain about 5000 water films instead of the one water film between clean grains. While the rate of diffusion in this network of films cannot be expected to be 5000 times as rapid as in a single film, a considerable increase in the rate of diffusion and hence the rate of pressure solution should result. To obtain this increase in rate, the film does not necessarily have to be made up of clay. The only requirement is that the film be composed of a material of very small grain size compared with the size of the mineral grains. It must be porous, the interstices must be saturated with water, and the material must itself not be subject to rapid pressure solution.

Our model therefore indicates that the 'catalytic' effect of clay postulated by Heald can be explained in terms of a physical enhancement of the rate of diffusion out of the contact zone. On the same basis, the formation of stylolites can be elucidated.

If initially the sediment contains a layer of clay, pressure solution will concentrate along this layer. Irregularities in grain size, mineralogy, and crystallographic direction will cause the amount of solution above and below the seam to be irregular, so that the initially smooth layer will be distorted into the typical stylolitic form. The average surface of the stylolite, however, will in general parallel the surface of the initial clay layer. The displacement or axis of the columns will be in the direction of the maximum principal stress, since intergranular surfaces perpendicular to the maximum principal stress will experience the largest stress gradients and hence will lead to the most rapid rate of pressure solution.

The process is automatically stabilized to produce a uniform displacement in the direction of maximum principal stress. If a particular segment should lag behind, the stress on it would increase and more rapid pressure solution would result. Too rapid a local rate, on the other hand, would relax the stress and thus lead to a decreased rate. The sum of the rates of solution above and below the seam must therefore be constant to result in a parallel displacement in the direction of the maximum principal stress. In tectonically relaxed areas, the maximum principal stress being vertical, the columns of the stylolite will be vertical, provided no tilting has taken place after formation of the stylolites. This simple physical model thus indicates that a pre-existing clay film may be the cause of stylolite formation. During the growth of the stylolite, the initial clay film will grow in thickness owing to addition of the insoluble residue.

The present theory only explains how a pre-existing clay layer will develop into a stylolite and is not concerned with the origin of the clay film. The simplest case occurs if the film is a depositional feature. In this case, the average surface of the stylolite must be a bedding plane. This supposition is confirmed by the fact that most stylolites parallel the bedding. In the case of transverse macrostylolites, the clay film must have been introduced after sedimentation. This may be the result of water flowing through fractures in the rock and either washing in clay or dissolving the walls of the fracture and leav-

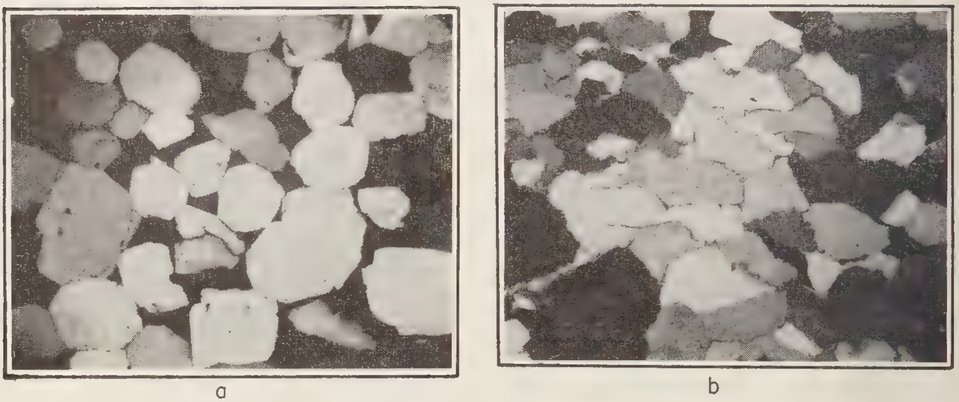


FIG. 18—Effect of clay on pressure solution. St. Peter sandstone, St. Charles Co., Missouri. Crossed nicols, ×55. (After *Heald* [1956]—reprinted by permission of the *Journal of Geology* and the University of Chicago Press.) (*a*) Clay-free sandstone. (*b*) Clay-bearing sandstone collected two inches below *a*.

ing them coated with the insoluble residue. Upon application of stress across the fracture, pressure solution will take place, and the coated fracture will develop into a stylolite.

The presence of the clay film has made it possible for the material tȯ be removed more rapidly from the zone of contact of the mineral grains. The dissolved mineral matter will enter the interstitial solution essentially in equilibrium with the exposed surfaces of the grains. Reprecipitation will take place on any favorable surfaces to which the solute will be transported by diffusion. The resultant compaction of the rock will be accompanied by an outflow of water which may remove some of the dissolved material or may bring in additional cementing material, depending on the concentration gradients and the direction of flow. The stylolitic seam, since it is made up of fine-grained material, will generally have a much lower permeability than the rock, so that the direction of fluid flow will tend to parallel the average surface of the stylolite. The flow, however, is between the stylolites and not along the distorted stylolitic seam. The rate of diffusion of solute in the seam depends primarily on the porosity and for similar geometries is independent of the grain size. The permeability, on the other hand, is very sensitive to changes in grain size and will be extremely low in the fine-grained material of the stylolitic seam.

It may be argued that clay films are always present between the mineral grains of a clastic sediment and that our ideal picture of a solvent film between clean mineral grains is never applicable. If this is so, one would expect that the product *Dmb* calculated for various 'clean' sands could vary widely owing to differences in this fine clay film. A fairly constant value of *Dmb* on the other hand would indicate either that we are dealing with clean grains or, less likely, that the amount of clay in the contact zone is constant for clean sands. Considerably more work in applying the theory to actual cases will have to be done before this question can be answered.

If the clay film becomes thick compared with the grain size, it will have the effect of averaging out the stresses across the clay-grain contact, so that the mineral grains are essentially under a uniform overburden pressure. This reduction in stress gradients will result in a reduced rate of pressure solution. Thus, while small amounts of clay will enhance the rate of pressure solution, large amounts of clay will have the opposite effect.

The effect of clay on pressure solution is illustrated in Figures 18*a* and 18*b*, which are Plates 1*D* and 1*C* from *Heald's* paper [1956]. His captions read: "*D* [our *a*], clay-free sandstone collected 2 inches above specimen 23 shown in *C* [our *b*]. *C*, highly pressolved, clay-bearing sandstone. Note sutured contacts, small amount of pore space, and elongation of grains parallel to bedding (horizontal). Compare with clay-free sandstone shown in *D*."

The Force of Crystallization

Whereas in pressure solution we were dealing with an outward transport of the solute in the solution film, the force of crystallization is a result of an inward diffusion of solute and a consequent growth of the mineral grains. If, as is usual, the minerals have a positive pressure coefficient of solubility, the force of crystallization can manifest itself only if the external solution is supersaturated. This supersaturation must exist in spite of the fact that precipitation may take place on the lateral faces. It may arise either from the slow rate of precipitation on these faces or from a greater solubility of these lateral faces due to a different crystallographic orientation. The boundary condition at the periphery of the cylinders (Fig. 1) is therefore determined by the supersaturation $\triangle c$ and the stress coefficient of solubility b according to the following linear approximation:

$$c = c_s + \Delta c = c_s + b\sigma(R) \qquad (28)$$

where c_s is the concentration of the saturated solution under the particular external hydrostatic pressure. The general solution gave four regions in which the effective normal stress at the circumference $\sigma(R)$ was positive (Fig. 4). Three of these—IV, V, and VI—are associated with a positive stress gradient and hence result in an inward flux of solute, whereas region I has a negative gradient and hence is a special example of pressure solution. In discussing these solutions, we shall first consider model A and then the mathematically more complex model B.

Model A—For Model A the effective normal stress across the solution film is given by equation (8a):

$$\sigma(x) = A_1x^2 + A_2$$

For a supersaturation $\triangle c$, the stress at the outer boundary ($x = 1$) must be

$$\sigma(1) = A_1 + A_2 = \Delta c/b \qquad (29)$$

Eliminating A_2, we obtain

$$\sigma(x) = A_1(x^2 - 1) + \Delta c/b \qquad (30)$$

Physically, the value of the integration constant A_1 is determined by the average effective normal stress $\bar{\sigma}$. In regions I, IV, and V, $\bar{\sigma}$ is obtained by integrating $\sigma(x)$ over the contact area and dividing by the area of the unit circle π. In region VI, where the stress goes to zero at some value of x between zero and one, the integration must be restricted to the part of the contact area where the stress is positive. Let X be this value of x:

$$\sigma(X) = 0 = A_1(X^2 - 1) + \Delta c/b \qquad (31)$$

Therefore

$$X = \sqrt{1 - \frac{\Delta c}{A_1 b}} \qquad (32)$$

In region V, which is the boundary between regions IV and VI, $X = 0$; therefore

$$A_1 = \Delta c/b \quad \text{(Region V)} \qquad (33)$$

Averaging the effective normal stress over the surface, we obtain

$$\bar{\sigma} = \frac{1}{\pi}\int_0^1 2\pi x \frac{\Delta c}{b} x^2\, dx = \frac{\Delta c}{2b} \quad \text{(Region V)} \qquad (34)$$

If the average effective stress is less than $\triangle c/2b$, we are in region VI, and growth will take place around the periphery, leaving a hollow center of fractional radius X. In this case the average effective stress is

$$\bar{\sigma} = \frac{1}{\pi}\int_X^1 2\pi x\left[A_1(x^2 - 1) + \frac{\Delta c}{b}\right] dx = \frac{\Delta c^2}{2A_1b^2} \quad \text{(Region VI)} \qquad (35)$$

If the average effective stress is larger than $\triangle c/2b$, we are in region I or IV, and the average effective stress is

$$\bar{\sigma} = \frac{1}{\pi}\int_0^1 2\pi x\left[A_1(x^2 - 1) + \frac{\Delta c}{b}\right] dx = \frac{\Delta c}{b} - \frac{A_1}{2} \quad \text{(Regions I, IV)} \qquad (36)$$

The results for the various regions as well as the rates of pressure growth or solution obtained by substituting for A_1 in terms of $\bar{\sigma}$ in equation (11a) are summarized in Table 2.

Model B—For Model B the effective normal

Table 2

PRESSURE SOLUTION OR CRYSTALLIZATION FROM SOLUTION OF SUPERSATURATION Δc

REGION		PRESSURE SOLUTION		FORCE OF CRYSTALLIZATION		
		NO SUPERSATURATION II	III	IV	V	WITH HOLLOW CENTER VI
Effective Normal Stress, σ		$\sigma(0) > \sigma(1) = 0$	$\sigma(0) > \sigma(1) > 0$	$\sigma(1) > \sigma(0) > 0$	$\sigma(1) > \sigma(0) = 0$	$\sigma(1) > \sigma(X) = 0, \quad 0 < X < 1$
[illegible]	Average Effective Normal Stress, $\bar\sigma$	$-\frac{A_1}{2}$	$\frac{\Delta c}{b} - \frac{A_1}{2}$	$\frac{\Delta c}{b} - \frac{A_1}{2}$	$\frac{\Delta c}{2b}$	$\frac{\Delta c^2}{2A_1 b^2}$
	Limits		$\bar\sigma > \frac{\Delta c}{b}$	$\frac{\Delta c}{b} > \bar\sigma > \frac{\Delta c}{2b}$		$\bar\sigma < \frac{\Delta c}{2b}$
	Rate of Growth, dr/dt	$-\frac{8Dmb}{R^2}\bar\sigma$	$-\frac{8Dmb}{R^2}\left[\bar\sigma - \frac{\Delta c}{b}\right]$	$\frac{8Dmb}{R^2}\left[\frac{\Delta c}{b} - \bar\sigma\right]$	$\frac{4Dm\Delta c}{R^2}$	$\frac{2Dm\Delta c^2}{R^2 b\bar\sigma}$
	Fractional Radius of Hollow Center, X				$X = 0$	$X = \left[1 - \frac{2b\bar\sigma}{\Delta c}\right]$
[illegible]	Average Effective Normal Stress, $\bar\sigma$	$\sigma^*\left[1 - (B_1 - 1)\ln\left[1 - \frac{1}{B_1}\right]\right]$	$\sigma^*\left[1 + xB_1e^{-x} + (B_1e^{-x} - 1)\ln\left[e^{-x} - \frac{1}{B_1}\right]\right]$		$\sigma^*\left[1 + \frac{x}{1 - e^{x}}\right]$	$\sigma^* B_1[1 - e^{-x}(1 + x)]$
	Limits	$\sigma^* > \bar\sigma > 0$	$\sigma^*(x + 1) > \bar\sigma > \sigma^* x + \frac{\Delta c}{b}$	$\frac{\Delta c}{b} > \bar\sigma > \sigma^*\left[1 + \frac{x}{1 - e^{x}}\right]$	$\frac{\Delta c}{2b} - \frac{\Delta c^2}{12b^2\sigma^*} + \cdots$	$\bar\sigma < \sigma^*\left[1 + \frac{x}{1 - e^{x}}\right]$
	Rate of Growth, dr/dt	$\frac{8Dmb\sigma^*}{R^2 B_1}$			$\frac{4Dmb\sigma^*}{R^2}(1 - e^{-x})$	$\frac{4Dmb}{R^2}\frac{\sigma^*}{x}[1 - e^{-x}(1 + x)]$
		$B_1 > 1$	$B_1 > e^{+x}$	$B_1 > \frac{1}{1 - e^{-x}}$	$\frac{4Dm\Delta c}{R^2}\left[1 - \frac{\Delta c}{24b\sigma^*} + \cdots\right]$	$\frac{2mD\Delta c^2}{R^2 b\bar\sigma}\left[1 - \frac{2}{3}\frac{\Delta c}{b\sigma^*} + \cdots\right]$
	Fractional Radius of Hollow Center, X				$X = 0$	$X = [1 + B_1(1 - e^{-x})]$

A_1, B_1 = Integration constants
b = Stress coefficient of solubility
Δc = Supersaturation
D = Diffusion constant

e = Base of natural logarithm = 2.718
h = Thickness of solution film
R = Radius of cylinders
t = Time

stress across the solution film is given by equation (8b):

$$\sigma(x) = \sigma^* \ln \frac{B_1}{x^2 + B_2}$$

For a supersaturation $\triangle c$, the stress at the outer boundary ($x = 1$) must be

$$\sigma(1) = \sigma^* \ln \frac{B_1}{1 + B_2} = \frac{\Delta c}{b} \qquad (37)$$

Eliminating B_2 and letting $\Delta c/\sigma^* b = \alpha$, we obtain

$$\sigma(x) = -\sigma^* \ln \left[\frac{x^2 - 1}{B_1} + e^{-\alpha}\right] \qquad (38)$$

If X is the value of x at which σ is zero, we obtain

$$X = \sqrt{1 + B_1(1 - e^{-\alpha})} \qquad (39)$$

In region V, $X = 0$. Therefore

$$B_1 = -\frac{1}{1 - e^{-\alpha}} \quad \text{(Region V)} \qquad (40)$$

Averaging the effective normal stress over the surface, we obtain

$$\bar{\sigma} = \frac{\sigma^*}{\pi} \int_0^1 2\pi x \ln \frac{1}{1 - x^2(1 - e^{-\alpha})}\, dx$$
$$= \sigma^* \left[1 + \frac{\alpha}{1 - e^{+\alpha}}\right] \quad \text{(Region V)} \qquad (41)$$

If the average effective stress is less than σ^* $[1 + \alpha/(1 - e^{+\alpha})]$, we are in region VI, and growth will take place around the periphery, leaving a hollow center of fractional radius X. In this case, the average effective stress is

$$\bar{\sigma} = -\frac{1}{\pi} \int_X^1 2\pi x \sigma^* \ln \left[\frac{x^2 - 1}{B_1} + e^{-\alpha}\right] dx$$
$$= -\sigma^* B_1[1 - e^{-\alpha}(1 + \alpha)] \quad \text{(Region VI)} \qquad (42)$$

If the average effective stress is larger than $\sigma^*[1 + a/(1 - e^{-a})]$, we are in region I or IV, and the average effective stress is

$$\bar{\sigma} = -\frac{1}{\pi} \int_0^1 2\pi x \sigma^* \ln \left[\frac{x^2 - 1}{B_1} + e^{-\alpha}\right] dx$$
$$= \sigma^* \Bigg[1 + \alpha B_1 e^{-\alpha} + (B_1 e^{-\alpha} - 1)$$
$$\cdot \ln \left(e^{-\alpha} - \frac{1}{B_1}\right)\Bigg] \quad \text{(Region I, IV)} \qquad (43)$$

For model B it is not possible for all regions to solve explicitly for the integration constant B_1 in terms of the average effective stress $\bar{\sigma}$. Rather, knowing $\bar{\sigma}$, we must evaluate B_1 for each particular case. Knowing B_1, we can then evaluate the rate of crystal growth or pressure solution by using equation (11b). The various results are summarized in Table 2.

Discussion—In discussing the case of force of crystallization, we shall consider what happens if we have two cylinders under stress and gradually increase the average effective normal stress. In order to be able to treat the example numerically, we shall assume that the ratio of the supersaturation to the pressure coefficient of solubility is unity with dimensions of a reciprocal stress and that the stress constant of the film for model B is one unit of stress. Thus the coefficient $\alpha = \triangle c/b\sigma^*$ is also unity.

As we start applying a load on the cylinders, we will be in region VI, and the force of crystallization will manifest itself by growth with a hollow center. At zero stress the rate of growth should be infinite for the assumptions here made. In an actual experiment, the rate will be limited by the rate at which solute can diffuse to the outer boundary of the cylinders. As the stress is increased, the rate of growth will decrease inversely to the stress, and the radius of the hollow center will shrink (Fig. 19).

The phenomenon of growth with a hollow center has been repeatedly demonstrated in the experiments of *Becker and Day* [1916] and *Taber* [1916]. They grew crystals of alum in an evaporating solution. As the water evaporated, the solution became supersaturated and crystal growth took place. The crystals not only grew on the top and sides but also grew a rim around the bottom which lifted the growing crystal from its support. This lifting took place even if a substantial weight was placed on the crystal. Unfortunately the observations reported do not permit a quantitative comparison with the theory. To permit such a comparison, the amount of supersaturation would have had to

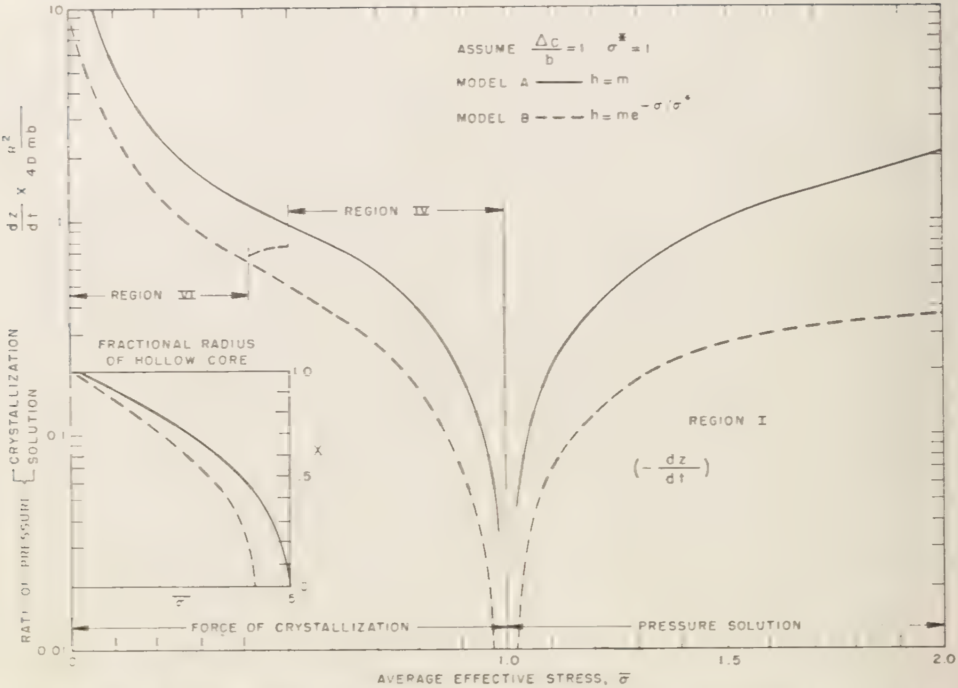


Fig. 19—Rate of pressure (crystallization solution) from supersaturated solution as a function of the average effective stress.

be carefully controlled. The experiments do, however, furnish a qualitative confirmation; thus when *Becker and Day* [1916, p. 324] placed a 0.7-gm weight on the growing crystal, the rate of lifting was decreased by a factor of 25 over that of a crystal stressed only by its own weight.

As the average effective stress is further increased, the rate of growth decreases, and the radius of the hollow center shrinks until finally it disappears. At this point we are in region V. For the assumptions of model A, this point is reached when the average effective stress is, for our numerical example, equal to one-half, whereas in the case of model B we reach this point at 0.418. In general, the point of transition will depend on the detailed mechanical properties of the solution film. The transition will take place earlier if the thickness of the film shrinks at a more rapid rate. This is shown by the series expansion for $\bar{\sigma}$ in region V.

$$\bar{\sigma} = \frac{\Delta c}{2b}\left(1 - \frac{\Delta c}{6b\sigma^*} + \cdots\right) \qquad (44)$$

If the film shrinks more rapidly, the characteristic stress of the film (σ^*) is smaller, and hence the average effective stress for region V is reduced.

As the average effective stress is further increased, the rate of crystal growth continues to decrease and goes to zero as the stress approaches unity. At this point the stress across the film is constant; hence there are no concentration gradients, and therefore there is no diffusive transfer. The transition occurs when the average effective stress is equal to the supersaturation divided by the pressure coefficient of solubility irrespective of the mechanical properties of the film.

A further increase of the average effective stress changes the direction of the solubility gradient, and pressure solution takes place with increasing rate as the average stress is increased. For model B, there exists a limiting average stress at which the film thickness at the center of the cylinder becomes zero. For higher stresses the film is broken and the proposed mechanism breaks down. In our example, this point is reached when the effective average stress becomes 2. This breaking of the film is a result of the specific assumption of a film whose thickness decreases exponentially as the stress is increased. A slight modification of this, such as adding a constant thin layer of water (for example, a monoatomic film) to the exponentially decreasing part, will get rid of this limitation. In nature, although there may be a limiting stress above which we get partial mineral-to-mineral contact and hence a breaking of the solution film, we should not expect the value of this stress to correspond to our simplified model.

Geologically the important problem is under what conditions we may expect the phenomenon of the force of crystallization. As pointed out by *Boydell* [1926], the force of crystallization is definitely a limited one. The mechanism here proposed indicates that crystallization will stop and pressure solution will take over if the average effective normal stress across the solution film exceeds the ratio of the supersaturation to the stress coefficient of solubility. Therefore, it is not possible to place a simple limit on this force in terms of a specific amount of overburden above which the force of crystallization cannot take place. As the solubilities of minerals and hence the amount of supersaturation that might be expected vary over many orders of magnitude, it is more convenient to divide the numerator and denominator of our ratio by the concentration and to consider the ratio of the fractional supersaturation and the fractional change in solubility with stress.

In the case of calcite, data given by *Owen and Brinkley* [1941] permit a rough estimate of the fractional change of solubility with hydrostatic pressure. Their data indicate that the fractional change of solubility with pressure is of the order of 10^{-3} per atmosphere. Assuming that the pressure coefficient under hydrostatic conditions is the same as the stress coefficient in the solution film, we see that the force of crystallization will be effective for a 1 per cent supersaturation up to effective normal stresses of 10 atmospheres. Should the supersaturation be higher, the stresses against which the calcite crystals can grow will be proportionately higher.

Minerals having a smaller but positive fractional pressure coefficient of solubility will be able to exert a force of crystallization against higher stresses or with smaller supersaturation. It is not within the scope of the present paper to explore the possible values of this coefficient or to investigate how and how much supersaturation may be expected under geologic situations. Our aim is merely to indicate the empirical parameters that determine the rate and direction of pressure crystallization. Assuming that the mechanism here proposed corresponds to reality, if we invoke pressure solution to explain a particular geologic situation, we would have to demonstrate that the ratio of the probable fractional supersaturation to the fractional pressure coefficient of solubility for the particular mineral was less than the average effective stress. Further, we would have to demonstrate that the rate of the process predicted by the theory is of the correct order of magnitude. Owing to the approximate nature of the theory and the uncertainty of the geologic parameters, we cannot, of course, expect to get an accurate prediction; however, we should be able to show that the numbers we obtain are roughly correct. Thus, although it will not be possible to show whether the theory is quantitatively correct, it will be relatively easy to demonstrate whether the assumptions are qualitatively incorrect.

Miscellaneous

In our previous discussions we have covered five of the six regions of our general solution. We have omitted region III (Fig. 6), in which the effective normal stress goes to zero before the periphery of the cylinder is approached. This region corresponds to pressure solution in an undersaturated solution so that the outer region of the contact zone is dissolved by the external solution, while pressure solution is active over the rest of the contact zone.

In all of our discussions we have assumed that the stress coefficient of solubility is positive; that is, that the minerals are more soluble under stress than without stress. Whereas this is almost certainly the case for calcite and quartz, some minerals may exist which have a negative pressure coefficient of solubility. Such minerals will have a rather strange behavior, for at points of stress the force of crystallization will manifest itself and will further increase the stress. We are thus dealing with an unstable system which will result in the growth of needles in the direction of maximum principal stress. In this case, the effect of the force field on the diffusion current may, however, reverse the direction of diffusion (see Appendix).

Appendix

Diffusion under the influence of external forces—In the theoretical development we have assumed that the diffusion current J is proportional to the concentration gradient; that is,

$$J = -D\nabla c \tag{45}$$

This equation is valid only if there are no external forces acting on the diffusing ion. In the presence of external forces, an additional term proportional to the force times the mobility must be added. Using the Nernst-Einstein relation [*Jost,* 1952, p. 47, 139] to relate the diffusion constant to the mobility, we find that the total current is

$$J = -D\left(\nabla c + \frac{c}{RT}\nabla W\right) \tag{46}$$

where the concentration c is in moles per cubic centimeter and W is the potential energy per mole. If the concentration and the potential energy depend only on one variable, say the pressure p, equation (46) can be written

$$J = -\frac{Dc}{RT}\left[RT\frac{d\ln c}{dp} + \frac{dW}{dp}\right]\nabla p \tag{47}$$

We shall illustrate this relationship by first considering the diffusion of a solute in the presence of a gravitational field.

Diffusion in a gravitational field—Consider a tall beaker containing a saturated salt solution with a crystal of salt near the top and another one near the bottom. If the solubility of the salt increases with hydrostatic pressure, the concentration around the bottom crystal, where the pressure is higher, will be greater. If we accept the fact that the diffusion current is proportional to the concentration gradient (equation 45), we obtain an upward diffusion, and the top crystal should grow at the expense of the bottom crystal. If the salt is denser than the solution, this will lead to an increase in potential energy and hence to a perpetual-motion machine of the second kind. We shall next show that the use of the correct expression (equation 47) leads to a downward diffusion, in opposition to the concentration gradient, resulting in a decrease in potential energy.

Let M be the molecular weight of the salt, d the density of the solid, and d_s and d_w the densities of the solution and the solvent, respectively. Let v be the partial molar volume of the solute in the solution. For a dilute solution, the isothermal change of solubility with pressure is given by

$$RT\frac{d\ln c}{dp} = \frac{M}{d} - v \tag{48}$$

The molar change in potential energy with pressure is

$$\frac{dW}{dp} = -\frac{M - vd_w}{d_s} = \frac{dW}{dz}\frac{dz}{dp} \tag{49}$$

Substituting (48) and (49) into (47) gives

$$J = \frac{Dc}{RT}\left[M\left(\frac{1}{d_s} - \frac{1}{d}\right) + v\left(1 - \frac{d_w}{d_s}\right)\right]\nabla p \tag{50}$$

Since $d > d_s$ and $d_s > d_w$, the right-hand side of (50) must be positive, and hence the diffusion current is in the direction of increasing pressure (downwards), although the concentration increases in a downward direction.

The case of diffusion in a gravitational field thus illustrates the importance of the transport due to external forces. Neglect of this term not only alters the magnitude but also the sign of the diffusion current. Having considered the relatively simple case of diffusion in a gravitational field, we shall next consider the diffusion current in a stressed solvent film.

Diffusion in a stressed solvent film—In the stressed film, we obtain a contribution to the potential energy due to the difference in apparent compressibility between the solute (K) and the solvent (K_w). These apparent compressibilities are defined as

$$K = -\frac{1}{v}\frac{dv}{dp}\bigg|_T, \quad K_w = -\frac{1}{v_w}\frac{dv_w}{dp}\bigg|_T \tag{51}$$

where v and v_w are the partial molar volumes of the solute and the solvent, respectively. The molar potential energy is

$$W = \frac{v}{2}(K - K_w)p^2 \tag{52}$$

Assuming that (48) applies for the change in solubility with normal stress, we obtain for the diffusion current

$$J = -\frac{Dc}{RT}\left[\frac{M}{d} - v + v(K - K_w)p\right]\nabla p \tag{53}$$

The term $M/d—v$ is the negative change in partial molar volume on solution of the solute. If this term is positive it indicates that the solvent has been compressed around the solute. As a result of this compression the apparent compressibility of the solute will be smaller than that of the solvent. If, on the other hand, the partial molar volume of the solute increases on solution, the solvent around the solute is expanded, resulting in a greater apparent compressibility of the solute relative to the solvent. In either case the effect of the stress energy term is to reduce the diffusion current.

This greatly oversimplified analysis gives an indication of the effect of the stress energy on the diffusional transport. A quantitative estimate, however, will require a more detailed knowledge of the properties of the assumed films.

Acknowledgments—The present paper resulted from a stimulating discussion with M. King Hubbert on the force of crystallization. I am especially grateful to R. C. Vernon, Shell Oil Company, for supplying me with data on the geologic history of the St. Peter sandstone, to R. J. Dunham for supplying the photomicrographs of the crinoidal limestone, to M. T. Heald for permission to use the photomicrographs from his paper on the St. Peter, and to A. S. Ginzbarg for the numerical computations. In addition, the author has had many helpful discussions with J. T. Smith, G. Rittenhouse, H. Rainbow, H. Odé, and R. L. Chuoke.

References

Becker, G. F., and A. L. Day, Note on the linear force of growing crystals, *J. Geol., 24,* 313–333, 1916.

Boydell, H. C., Metasomatism and linear "force of growing crystals," *Econ. Geol., 21,* 1–55, 1926.

Heald, M. T., Cementation of Simpson and St. Peter sandstones in parts of Oklahoma, Arkansas, and Missouri, *J. Geol., 64,* 16–30, 1956.

Henniker, J. C., Depth of the surface zone of a liquid, *Revs. Modern Phys., 21,* 322–342, 1949.

Jost, W., *Diffusion in Solids, Liquids, and Gases,* Academic Press, New York, 558 pp., 1952.

Kennedy, G. C., A portion of the system silica-water, *Econ. Geol., 45,* 629–653, 1950.

Owen, B. B., and S. R. Brinkley, Calculation of the effect of pressure upon ionic equilibrium in pure water and in salt solutions, *Chem. Revs., 29,* 461–474, 1941.

Stockdale, P. B., Stylolites—their nature and origin, *Indiana Univ. Studies, 9,* no. 55, 1–97, 1922.

Taber, S., The growth of crystals under external pressure, *Am. J. Sci., 41,* 532–556, 1916.

Weyl, P. K., The solution kinetics of calcite, *J. Geol., 66,* 163–176, 1958.

(Manuscript received June 3, 1959.)

Letters to the Editor

MAGNETIC CUTOFF RIGIDITIES OF CHARGED PARTICLES IN THE EARTH'S FIELD AT TIMES OF MAGNETIC STORMS

P. ROTHWELL

State University of Iowa, Iowa City, Iowa

Recent observations from balloons and satellites have indicated that the low-energy solar cosmic rays that quite frequently bombard the upper atmosphere at high latitudes for several days after a large solar flare [*Anderson and others,* 1959] can arrive at a given location with energies below the usual magnetic cutoff energy after the onset of the magnetic storm associated with the flare [*Freier, Ney, and Winckler,* 1959; *Ney, Winckler, and Freier,* 1959; *Rothwell and McIlwain,* 1959]. The decrease in the intensity of galactic cosmic rays during magnetic storms, the well known Forbush effect, generally obscures any increase in sea-level cosmic-ray intensity that might be expected from an additional flux of low-energy cosmic rays and lowered magnetic cutoff energies. *Yoshida and Wada* [1959], however, have recently reported small sea-level increases in cosmic-ray intensity superimposed on the Forbush decreases which occurred at the time of the very large magnetic storms on September 13, 1957, and February 11, 1958, when auroras were observed as far south as Tokyo.

It is possible to calculate the changes in the Störmer magnetic cutoff rigidities for charged particles in the earth's field at the time of magnetic storms by supposing that: (1) a magnetic storm cloud of solar gas can push into the earth's field to within x earth radii of the dipole origin in the equatorial plane (the larger the storm, the smaller the value of x); (2) particles can move into the earth's field at x earth radii via the storm cloud; (3) the earth's field is little affected by the storm cloud at distances less than x earth radii from the dipole origin.

Störmer theory shows that charged particles can move in a dipole field only within two 'allowed' zones. The boundaries of these zones are the toroidal surfaces with meridian curves given by the equations

$$\left.\begin{aligned} r &= \frac{\cos^2\psi}{\gamma + \sqrt{\gamma^2 + \cos^3\psi}} \qquad (1)\\ r &= \frac{\cos^2\psi}{\gamma + \sqrt{\gamma^2 - \cos^3\psi}} \qquad (2)\end{aligned}\right\}\ \text{inner allowed zone}$$

$$r = \frac{\gamma + \sqrt{\gamma^2 - \cos^3\psi}}{\cos\psi} \quad \text{outer allowed zone} \qquad (3)$$

where γ is a constant proportional to the z component of angular momentum about the origin of a particle at infinity, ψ is measured from the equator, and r is expressed in Störmer units. The Störmer unit C_{st} is given by the expression

$$C_{st} = \sqrt{(3/10^7)\times(M/P)} \quad \text{cm}$$

where M is the moment of the dipole (in gauss cm^3) and P the rigidity of the particle (in Bv). For all values of γ greater than 1, the two 'allowed' zones are separated by a 'forbidden' region and are mutually inaccessible. Particles can reach the origin from infinity only when the inner and outer allowed zones connect, that is, when γ is less than or equal to 1. The two zones just connect when $\gamma = 1$, and particles can just move into the inner allowed zone from outside at $r = 1$ stormer in the equatorial plane. If the radius of the earth is r_e stormers, the lowest latitude that a particle can reach on the earth's surface is given by equation 1, with $\gamma = 1$ and $r = r_e$; that is,

$$\cos^2\psi = r_e(1 + \sqrt{1 + \cos^3\psi}) \qquad (4)$$

$$\cos\psi \simeq \sqrt{2r_e} \qquad (5)$$

when $\cos^3\psi \ll 1$.

The cutoff rigidity P_c for a vertical charged particle arriving at a given latitude λ is obtained from equation 5 by converting the radius of the earth from stormers into metric units; that is,

$$P_c = (3/10^7)(M/4R_e^2)\cos^4\lambda \qquad (6)$$

where R_e = radius of the earth = 6.4×10^8 cm. When a magnetic storm cloud pushes into the earth's field, to within x earth radii of the dipole origin, charged particles move into the inner allowed zone at $r = xr_e$ stormers in the equatorial plane. The maximum value of γ for which the inner allowed zone would be accessible to particles from outside (when xr_e is less than 1 stormer), is obtained from equation 2, and may be written

$$\gamma_{max} = 1/2[xr_e + (1/xr_e)] \qquad (7)$$

The lowest latitude that a particle can reach on the earth's surface in this case is given by the expression

$$\cos\psi = \sqrt{(1/x) + xr_e^{\,2}} \qquad (8)$$

and the 'cutoff' rigidity for a particle arriving vertically at latitude λ is

$$P_c = (3/10^7)(M/R_e^2)(1/x)[\cos^2\lambda - (1/x)] \; (9)$$

when $xr_e < 1$. It follows from this equation that, when $\cos^3\lambda \leq 1/x$, particles of all rigidities should be able to reach the earth's surface.

Remembering that the equation of a line of force in a dipole field is $r_e = r_0 \cos^2\lambda$ (where r_0 is the distance, in stormers, from the dipole origin, at which the dipole line cuts the equatorial plane), so that

$$r_e/r_0 = R_e/R_0 = \cos^2\lambda$$

then the conditions under which equations 6 and 9 apply, $xr_e \geq 1$, $xr_e < 1$, can be written $x \geq 2R_0/R_e$ and $x < 2R_0/R_e$, respectively, and the condition $\cos^3\lambda \leq 1/x$ can be written $x \leq R_0/R_e$.

To summarize: the cutoff rigidities for charged particles at a latitude λ, when a magnetic storm cloud pushes into the earth's field to within x earth radii of the dipole origin, are given by the following expressions

$$P_c = \frac{3}{10^7}\cdot\frac{M}{4R_e^{\,2}}\cos^4\lambda$$

$$\text{when} \quad x \geq 2\frac{R_0}{R_e}$$

$$P_c = \frac{3}{10^7}\cdot\frac{M}{R_e^{\,2}}\cdot\frac{1}{x}\left(\cos^2\lambda - \frac{1}{x}\right)$$

$$\text{when} \quad 2\frac{R_0}{R_e} > x > \frac{R_0}{R_e}$$

$$P_c = 0$$

$$\text{when} \quad x \leq \frac{R_0}{R_e}$$

TABLE 1

Place	λ_{dip}	R_0		$x=\infty$	$x=10$	$x=7$	$x=5$	$x=4$	$x=3$	$x=2.5$	$x=2.0$	$x=1.5$	$x=1.2$
Churchill	78°	$23R_e$	P_c,Bv	0.03	0	0	0	0	0	0	0	0	0
			E_p,bev	5×10^{-4}	0	0	0	0	0	0	0	0	0
College Fairbanks	64°	$5.2R_e$	P_c,Bv	0.51	0.49	0.38	0	0	0	0	0	0	0
			E_p,bev	0.12_5	0.11_5	0.07	0	0	0	0	0	0	0
Minneapolis	61°	$4.2R_e$	P_c,Bv	0.85	0.85	0.81	0.42	0	0	0	0	0	0
			E_p,bev	0.32	0.32	0.29	0.08_5	0	0	0	0	0	0
Iowa City	55°	$3.1R_e$	P_c,Bv	1.65	1.65	1.65	1.56	1.18	0	0	0	0	0
			E_p,bev	0.95	0.95	0.95	0.88	0.66	0	0	0	0	0
Bristol	50°	$2.4R_e$	P_c,Bv	2.55	2.55	2.55	2.55	2.45	1.60	0.30	0	0	0
			E_p,bev	1.75	1.75	1.75	1.75	1.65	0.90	0.04_5	0	0	0
San Angelo	42°	$1.8R_e$	P_c,Bv	4.8	4.8	4.8	4.8	4.8	4.6	3.7	1.6	0	0
			E_p,bev	3.8	3.8	3.8	3.8	3.8	3.6	2.7	0.9	0	0
Tokyo	30°	$1.3R_e$	P_c,Bv	9.7	9.7	9.7	9.7	9.7	9.7	9.6	8.6	3.7	0
			E_p,bev	8.8	8.8	8.8	8.8	8.8	8.8	8.7	7.7	2.7	0
Huancayo	1°	$1.0R_e$	P_c,Bv	14.2	14.2	14.2	14.2	14.2	14.2	14.2	14.2	12.6	7.8
			E_p,bev	13.3	13.3	13.3	13.3	13.3	13.3	13.3	13.3	11.7	6.9

In reality, particles with very low rigidities would not move readily from the storm cloud into the earth's field, and the strength of the field in the cloud itself would probably set a lower limit to the rigidity of charged particles that can reach the earth's surface. Table 1 gives the cutoff rigidities for charged particles at various places, for different values of x, and the corresponding cutoff energies for protons. The values of M and λ are calculated from the local values of horizontal magnetic intensity H and dip angle δ, at the place considered, through the dipole relations

$$2 \tan \lambda_{dip} = \tan \delta$$

$$H = (M/R_e^3) \cos \lambda_{dip}$$

It has been shown that better agreement with experimentally measured cutoff rigidities is obtained by using the 'local' dipole magnetic moment and latitude in the Störmer equation than by using the moment and latitude of the single dipole which gives the best fit to magnetic observations over the whole of the earth's surface [*Rothwell,* 1958].

It can be seen from Table 1 that, as a storm cloud pushes into the earth's field, it should first lower the cutoff rigidity at a given latitude, and then finally remove it, at which point it would be reasonable to expect the appearance of aurora. For instance, if a storm cloud is composed of ionized hydrogen, moves with a velocity of 10^8 cm/sec, and contains about 1000 particles/cc, from energy considerations it should be able to push into the earth's field to about 3 earth radii in the equatorial plane, and so produce overhead aurora at magnetic latitudes of 55° and modify cutoff rigidities down to about 40°. A storm cloud moving with a velocity of 3×10^8 cm/sec and containing 10^4 particles/cc could modify cutoff rigidities even at equatorial latitudes. As the magnetic storm subsides, the solar cloud moves out of the earth's field, the cutoff rigidities return to their usual values, and some of the particles temporarily admitted to the Störmer inner allowed zone may remain trapped there, so that after a large storm additional particles may well be found even in the inner Van Allen radiation belt. It will be interesting to see how the intensity of the inner belt will vary during the declining solar cycle. An intensity decrease with decreasing solar activity would imply that solar injected particles contribute more to the intensity of the inner radiation belt than cosmic-ray albedo particles, because there are few large magnetic storms near solar minimum, and the cosmic-ray intensity is higher at solar minimum than solar maximum. On the other hand, an intensity increase with decreasing solar activity would still leave the question open since some magnetic storms may be more efficient at allowing trapped particles to escape than at injecting fresh particles into the trapping zone.

References

Anderson, K. A., R. Arnoldy, R. Hoffman, L. Peterson, and J. R. Winkler, Observations of low-energy solar cosmic rays from the flare of August 22, 1958, *J. Geophys. Research, 64,* 1133–1147, 1959.

Freier, P. S., E. P. Ney, and J. R. Winckler, Balloon observation of solar cosmic rays on March 26, 1958, *J. Geophys. Research, 64,* 685–688, 1959.

Ney, E. P., J. R. Winckler, and P. S. Freier, Protons from the sun on May 12, 1959, *Phys. Rev. Letters, 3,* 183, 1959.

Rothwell, P., Cosmic rays in the earth's magnetic field, *Phil. Mag.,* 3 (33), 961, 1958.

Rothwell, P., and C. E. McIlwain, Satellite observations of solar cosmic rays, *Nature, 184,* 138, 1959.

Yoshida, S., and M. Wada, Storm-time increase of cosmic-ray intensity, *Nature, 183,* 381, 1959.

(Received August 17, 1959)

DIRECTION FINDINGS ON WHISTLERS

J. M. Watts

National Bureau of Standards
Boulder, Colorado

Efforts to study the direction of arrival and polarization of whistlers have led to confusion, presumably because of the complexity of the signals and their transient behavior. This letter describes one technique which shows promise in that it has given results on the direction of some whistlers.

The system consists essentially of two loops in vertical planes at right angles connected to a goniometer driven by a scanning motor. The rotating loop thus simulated will exhibit nulls when it is broadside to signals with small angular spread if they arrive at low elevation angles. If the signal is not a point source the nulls will be shallow; and if it is downcoming and circularly polarized (as some whistlers are expected to be), the nulls should disappear. If the signals are downcoming and linearly polarized perpendicular to the plane of incidence, nulls will be detected, but they will be displaced in angle by 90° from the expected ones. The essential fact is that some whistlers do exhibit nulls when the goniometer is rotated. It is driven at a constant speed of 5 revolutions per second during the recording interval, and recorded whistlers are analyzed in the conventional way by means of a sound spectrograph. Since there is a 180° ambiguity in the bearing from a single location there is a possibility of 10 nulls per second. Bearings are estimated from a measuring scale constructed from the null patterns of the U. S. Navy VLF stations: NSS in Annapolis, Maryland, NPM in Hawaii, and NLK in Washington state, their azimuthal angles from the recording site being known.

Figure 1 illustrates the measurement of a whistler. Four nulls can be seen in the trace. The null pattern of station NSS is at the top. The whistler nulls have the same spacing, but are displaced by an angle equal to the bearing of the whistler relative to NSS. It is conceivable that some whistlers do not have a constant direction of arrival during their lifetime, or that they are composed of energy spread in direction. If they are not too complicated, these features can be studied by means of the simple apparatus described.

Direction finding in whistler research is important. Before the start of the IGY it was discovered that the same whistler could be heard at two places as much as 1000 miles apart; moreover, characteristic patterns of dispersion were apparently the same in both places. The implication was that the propagation paths contributing to the signals at both places were the same. At the meeting of the URSI in Washington, D. C., May 1959, R. A. Helliwell of Stanford University demonstrated convincingly that lightning flashes can excite both ends of a whistler path and that the path is unique as viewed from either end. These two observed characteristics point to direction finding as the most direct way of locating the path terminals. I believe that the large number of whistler recordings to be made during the next few years should be provided with this information. Nevertheless, the information from just one pair of stations, which will soon be available, should open a new phase in whistler research.

The use of natural whistlers for locating propagation ducts has a considerable advantage over the use of VLF station signals, namely its ability to separate several paths if they are present, provided that their dispersions and time delays allow their distinction. Also, if there is a change of bearing with frequency, which has not yet been noticed, the phenomenon can be studied. Whistlers of the "swishy" type, however, would not be expected to show direction by the null technique, since it is likely that their "swishiness" is caused by the contributions from a multiplicity of paths having different dispersions. This prediction has been borne out by the lack of definite nulls in the spectrograms of whistlers of that type.

(Received August 14, 1959.)

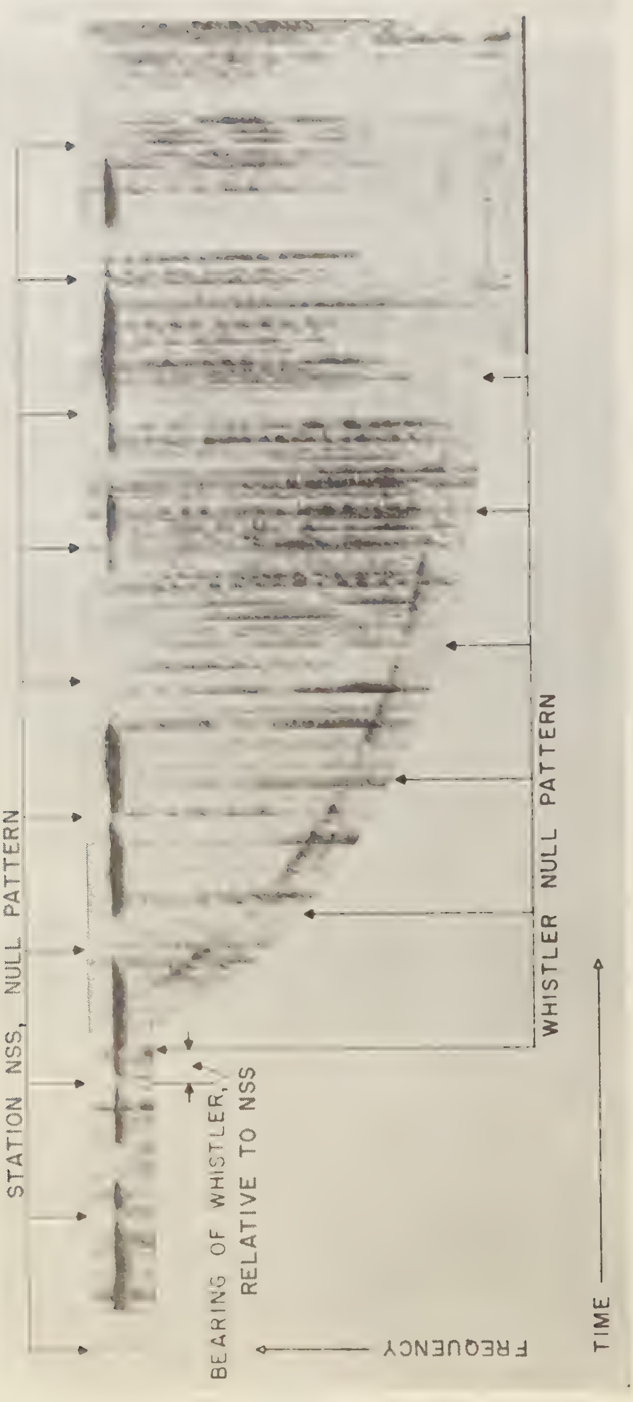


FIG. 1—Direction of a whistler.

A NOTE ON THE PAPER BY C. E. PALMER, 'THE STRATOSPHERIC POLAR VORTEX IN WINTER'

Jae R. Ballif

Institute of Geophysics
University of California
Los Angeles, California

Figures 4 and 6 of the paper [*Palmer*, 1959] illustrated the time variation of atmospheric pressure, temperature, and density at two stations in the core of the vortex prior to and during the warming of the arctic stratosphere in 1958. Since that time, the entire polar cap has been analyzed, using the complete IGY data at the 15-km level, to determine the daily distribution of these same variables. The map series from January 11 through January 30 is presently being prepared for publication.[1] The influence of the land and sea distribution is manifest in the asymmetry of the stationary features on the maps. Nevertheless, the synoptic study eradicates any doubt as to whether conditions above Alert or Clyde represent the entire region within the umbra.

To depict the changes, mean values of pressure, temperature, and density at various latitudes are plotted for three specified days (Fig. 1). These selected days bracketed January 24, when the local temperature changes were greatest. In the initial period, the pressure and density increased in a near isothermal process. During and after the warming the pressure continued to rise, the temperature warmed explosively, and the density dropped to approximately its original value. These changes occurred within the earth's shadow.

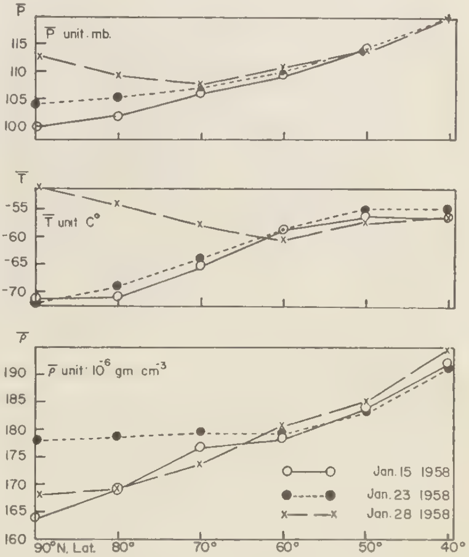


Fig. 1—Variation with latitude of pressure, temperature, and density at 15 km averaged over all longitudes for January 15, 23, and 28, 1958. Averaging has been carried out at 10° latitude intervals, beginning with the pole.

Reference

Palmer, C. E., The Stratospheric Polar Vortex in Winter, *J. Geophys. Research, 64,* 749–764, 1959.

[1] This work supported by the USNC for the IGY through the NSF and by ONR.

CONVERSION OF SEISMIC WAVES

J. N. Nanda

Office of Scientific Research and Development
Naval Headquarters
New Delhi, India

Tatel and Tuve [1954] have proposed that compressional waves impinging upon the earth's surface from below, at a point many kilometers distant from a shot were partially converted into Raleigh waves. Incidence of such converted energy explains the unrest after the arrival of the first compressional waves (allowance is made for the subsequent direct arrival of shear-Raleigh complex). It has been shown by the author [*Nanda,* to be published] that the decay of reverberation intensity in a record obtained close to a seismic shot cannot be explained on the basis of such a conversion into Raleigh waves but can be explained if the conversion is assumed to be into shear waves.

Tatel [1954] adduced experimental evidence in favor of conversion into Raleigh waves by means of a model experiment. The computed arrival time according to this hypothesis is 22.5 μ sec, whereas the observed time is 21.7 μ sec. It is, however, known that shear waves are 8.7 per cent faster than Raleigh waves and from the geometry of the experiment the difference is almost exactly accounted for if conversion into shear waves is assumed. The only other evidence in favor of Raleigh waves is from Tatel and Tuve's field experiments, but this is rather poor as Raleigh wave arrivals could be identified only in one out of three cases, and in this particular case interference of shear waves from some peculiar distribution of disturbing centers could be the cause of apparent Raleigh wave disturbance.

In this connection attention is invited to work of *Redwood* [1958], who has demonstrated in an exhaustive study of propagation of ultrasonic waves in solid cylinders the partial conversion into transverse waves when the compressional waves are incident on the bounding surface.

The conversion into S waves and the explanation of the reverberation decay in crustal seismology is important inasmuch as it will now become necessary to assume a fairly high rate of average increase of velocity with depth in the crust. This eventually will alter the usual computation of depth of penetration for high-intensity reflections (focusing effects) under land to a much smaller value. For example, the depth to the M discontinuity was determined by this method to be about 14 km, whereas a depth of 37 km is obtained according to the hypothesis of crustal layers of fairly uniform velocity.

References

Nanda, J. N., Evaluation of average crustal characteristics from reverberation of seismic waves (communicated for publication in the *Proc. Roy. Soc. London A*).

Redwood, M., The generation of secondary signals in the propagation of ultrasonic waves in bounded solids, *Proc. Phys. Soc., 72,* 841–853, 1958.

Tatel, H. E., Note on the nature of a seismogram —II *J. Geophys. Research, 59,* 289–294, 1954.

Tatel, H. E., and M. A. Tuve, Note on the nature of a Seismogram—I, *J. Geophys. Research, 59,* 287–288, 1954.

(Received July 17, 1959.)

SOME EXPERIMENTS IN POTASSIUM-ARGON DATING

Minoru Ozima

Department of Physics
University of Toronto
Toronto 5, Canada

In the usual potassium-argon dating method, potassium and argon analyses are each performed on a separate sample. However, this procedure sometimes leads to an erroneous age result because of the inhomogeneous distribution of potassium and argon in minerals. Hence, it is quite desirable to do both the potassium and argon analyses on the same aliquot of sample to eliminate this source of error. For this purpose, the distillation method, originally devised by *Edwards and Urey* [1955] to liberate alkalis from minerals has been modified so that both the potassium and argon are liberated from the same sample.

Two or three grams of the crushed mineral are put into an alundum crucible with half of this mass of calcium chloride. The crucible is suspended inside a pyrex glass container which is connected to a vacuum line. After evacuating the container, the tracer argon is released into the system, and the crucible is heated to 1350°C by an induction coil. Twenty minutes after starting to heat, the temperature reaches 1350°C. The crucible is maintained at this temperature for another 40 minutes. During this heating process the potassium in the mineral reacts with the calcium chloride to form potassium chloride, which evaporates from the crucible and condenses on the walls of the container. After all the evolved gases have been transferred into the purification line, the container is removed from the line for the potassium determination. The potassium chloride deposited on the glass wall is collected in a solution by washing the container wall with 1/5 normal hydrochloric acid. A subsequent analysis of this solution with a flame photometer gives the potassium content of the sample. The purified gas is collected in a break-seal tube for the argon determination.

This heat treatment is quite satisfactory for liberating both the potassium and the argon from minerals. After the heat treatment, a very small amount (about 2 or 3 per cent of the original sample) of transparent glass which remained in the crucible was examined. A second extraction experiment made on the residue showed no appreciable amount of argon and potassium.

In the determination of the radiogenic A^{40}, the sensitivity of a mass spectrometric analysis is greatly reduced by the presence of a small background of mass-36. A sweeping method in taking a mass spectrometric record freed us from the error due to the mass-36 background to give good sensitivity in determining the radiogenic A^{40}.

Immediately before introducing the sample gas into the analyzer tube of a mass spectrometer, the analyzer section (the analyzer tube, a diffusion pump, and a liquid air cold trap) is isolated from a main vacuum system. Then the gas introduced into the system is forced by the diffusion pump to recirculate the system. Since the sample gas is introduced through a very thin capillary tube, the pressure or the amount of the sample in the analyzer tube tends to increase during the recycling process, which lasts for about an hour. Figure 1 illustrates the mass spectrogram in a typical analysis. In treating the data, the peak heights of mass-38 and mass-40 are extrapolated to the intervening mass-36 peak from both sides. The average of the values extrapolated from the right-hand side [denoted by $A^{40}(+)$, $A^{38}(+)$ in Fig. (1)] and from the left-hand side [$A^{40}(-)$, $A^{38}(-)$] together with the average value of the corresponding mass-36 peak heights give one set of the isotopic ratios; that is, A^{36} to A^{38} and

Present address: Enrico Fermi Research Institute, University of Chicago, Chicago, Illinois.

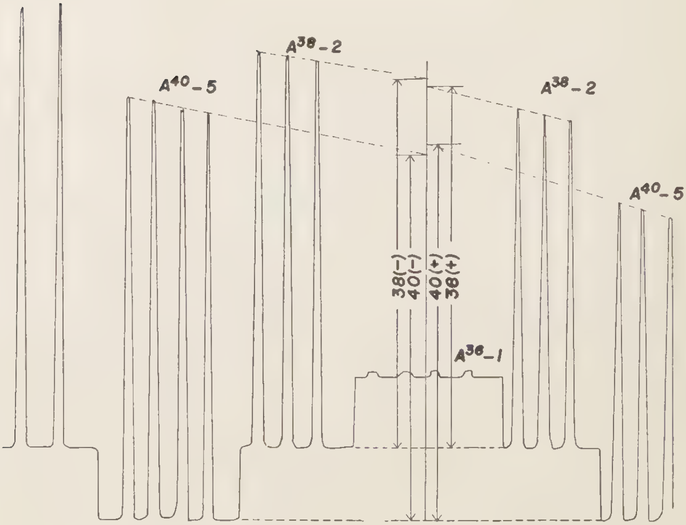


FIG. 1—Mass spectrogram in a typical analysis.

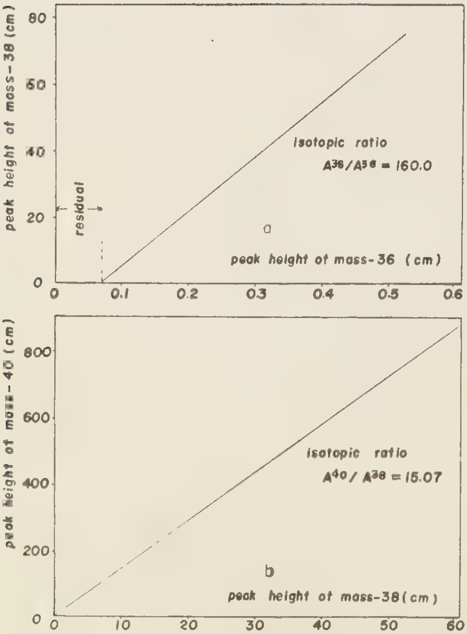


FIG. 2—Peak height ratios; (a) mass-38 with respect to mass-36; (b) mass-40 with respect to mass-38.

A^{38} to A^{40}. These values are plotted in Figure 2 with other values determined in the subsequent sweep of the mass spectrum. Figure 2a shows peak heights of a series of the mass-38 measurements with respect to the corresponding mass-36 peaks, and Figure 2b shows those of mass-40 to mass-38 for the same sample. Only argon gas is responsible for the variation of peak heights, as the impurities (mainly due to HCl^{+38} and hydrocarbons) which cause the background in mass-36 are trapped by a liquid air cold trap during the recycling process and these give a constant background level in a mass spectrogram. Consequently, the slopes of the lines give the ratios of A^{38} to A^{36} and A^{40} to A^{38} irrespective of the presence of the small background in the mass-36.

I am grateful to Dr. R. D. Russell for invaluable discussions.

REFERENCES

EDWARDS, G., and H. C. UREY, Determination of alkali metals in meteorites by a distillation process, *Geochim. et Cosmochim. Acta*, 7, 154–168, 1955.

(Received May 23, 1959; revised August 20, 1959.)

AUTHORS' REPLY TO DE VRIES AND PHILIP'S DISCUSSION OF 'EFFECT OF TEMPERATURE DISTRIBUTION ON MOISTURE FLOW IN POROUS MATERIALS'

W. Woodside[1] and J. M. Kuzmak[2]

Division of Building Research
National Research Council
Ottawa 2, Canada

We wish to reply to the comment of *de Vries and Philip* [1959] on our paper regarding the effect of temperature distribution on moisture flow in porous materials [*Woodside and Kuzmak,* 1958]. De Vries and Philip stated: 'In the case of heat transfer, the soil-air interface represents a surface of refraction of the stream surfaces. However, in the vapor field, this same interface is itself a stream surface, across which there can be no transport.' We do not agree that the above statement is applicable to the system which we simulated with our model.

We postulated the following:

1. The surface of the particle in our air-dry condition is covered with a film of water; therefore, the interface with air is *water-air,* not solid-air. (Note that at very low moisture contents, before the bone-dry condition is reached, when the interface is largely solid-air, the observed rate of flow agrees fairly closely with that predicted by Fick's law.)

2. Above the surface of the water film, if the influence of the curvature of the water surface is disregarded, there is water vapor at a vapor pressure which corresponds to the temperature of the water film at that point.

3. The water evaporates from the film surface and that this is the *source* of the vapor.

4. The water vapor distills from one liquid surface to that on an adjacent particle as a result of the vapor-pressure difference which corresponds to the temperature difference.

5. The evaporation of water in one area and the condensation of water in another area of a film results in a difference in film thickness which in turn results in a difference in suction between the thick and the thin parts of the film; the water in the thin portion of the film is at the lower free energy. The suction difference causes the water to move in the film phase from the thick to the thin film region, and this movement tends to restore the film to equilibrium thickness. If there is a film of water on the surface of a spherical particle, water condenses on one half of the total area and evaporates from the other half. The water flows because of the suction difference from the condensation area to the evaporation area in the film phase over the surface of the particle; it flows because of the vapor-pressure difference from the evaporation area of one particle to the condensation area of an adjacent particle in the vapor phase between the particles.

Thus the flow consists of multiple vapor and liquid flow in series. It is not merely vapor diffusion through the tortuous air paths of the porous medium. The primary driving force is the vapor-pressure difference which corresponds to the temperature difference. Flow in the film phase occurs because of the suction difference produced as a result of the flow in the vapor phase.

We measured the temperatures at certain points on the surface of the model and calculated the temperature gradients between mirror image points on the model. We further calculated an average value for the gradient. We neglected the 45 per cent volume or 25 per cent cross-sectional area referred to by de Vries and Philip because, for the model chosen, we assume that this region does not convey water vapor, having neither source of water nor sink in its path. Our calculated average temperature gra-

[1] Present address: Gulf Research and Development Company, Pittsburgh, Pa.

[2] Present address: American Viscose Corporation, Marcus Hook, Pa.

dient is six times the over-all gradient. This then could account for the discrepancy between the observed and the calculated flow rates.

It may be worth noting that experiments have been carried out which show that a suction difference applied across a porous material of a given moisture content results in flow entirely in the film phase. This is evidence that the film phase is continuous from one side of the porous material to the other. The data indicate that a temperature gradient, but no suction gradient, applied across this same porous material at the same moisture content does not result in continuous film flow from the hot to the cold side even though the film itself is continuous [*Kuzmak and Sereda,* 1957]

We would suggest that further experimental work might be done to check our observations and our hypothesis regarding the mechanism of flow. In the meantime, we do not accept de Vries and Philip's contention that our concept is incorrect.

References

DeVries, D. A., and J. R. Philip, Temperature distribution and moisture transfer in porous materials, *J. Geophys. Research, 64,* 386–388, 1959.

Kuzmak, J. M., and P. J. Sereda, The mechanism by which water moves through a porous material subjected to a temperature gradient, Part II, Salt tracer and streaming potential to detect flow in the liquid phase, *Soil Sci., 84,* 419–422, 1957.

Woodside, W., and J. M. Kuzmak, Effect of temperature distribution on moisture flow in porous materials, *Trans. Am. Geophys. Union, 39,* 676–680, 1958.

(Received July 28, 1959.)

Corrigendum

Dr. Philip Newman called attention to two errors in his paper, *Optical, electromagnetic, and satellite observations of high-altitude nuclear detonations, Part I,* page 927, August 1959 issue of this JOURNAL.

The reference on the eighth line of the first paragraph in the left-hand column should be [*Fowler and Waddington,* 1958] instead of [*Koomen et al.,* 1957]

The first two sentences of the first paragraph in the right-hand column should read as follows:

"A height of 100 km for the aurora at a range of 800 km requires an elevation angle of 3°20′, apparently below the visual aurora. In this case, the ground conjugate point is deviated by about 220 km to the northwest of the conjugate point for event III computed at the RAND Corporation by Vestine and Karzas."

Contents

(*Continued from back cover*)

(Journal of Geophysical Research, Vol. 64, No. 11, November 1959)

Contents

PAGE

SYMPOSIUM ON THE EXPLORATION OF SPACE

(Edited by Robert Jastrow)

(*Continued inside back cover*)